

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
7 juin 2001 (07.06.2001)

PCT

(10) Numéro de publication internationale
WO 01/40497 A2

(51) Classification internationale des brevets⁷: C12P 19/34,
C12N 15/52, 15/63, 15/11, 9/00, C12Q 1/68, C07K 16/40,
G01N 33/53, C12N 15/10, 15/52, 9/00, C12Q 1/68

(21) Numéro de la demande internationale:
PCT/FR00/03311

(22) Date de dépôt international:
27 novembre 2000 (27.11.2000)

(25) Langue de dépôt: français

(26) Langue de publication: français

(30) Données relatives à la priorité:
99/15032 29 novembre 1999 (29.11.1999) FR
60/209,800 7 juin 2000 (07.06.2000) US

(71) Déposant (pour tous les États désignés sauf US): AVEN-
TIS PHARMA S.A. [FR/FR]; 20, avenue Raymond Aron,
F-92160 Antony (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement): JEAN-
NIN, Pascale [FR/FR]; 52, rue Pierre Louvrier, F-92140
Clamart (FR). PERNODET, Jean-Luc [FR/FR]; 21,
rue des Jardins, F-94230 Cachan (FR). GUERINEAU,
Michel [FR/FR]; 79, boulevard Saint Marcel, F-75013
Paris (FR). SIMONET, Pascal [FR/FR]; 55, rue Pierre

Voyant, F-69100 Villeurbanne (FR). COURTOIS, Sophie
[FR/FR]; 165, rue de Paris, F-94220 Charenton le Pont
(FR). CAPPELLANO, Carmela [FR/FR]; 16, rue de
Neuilly, F-94120 Fontenay sous Bois (FR). FRANCOU,
François [FR/FR]; 76, boulevard de Lozère, F-91120
Palaiseau (FR). RAYNAL, Alain [FR/FR]; 52, avenue des
Tilleuls, F-91440 Bures sur Yvette (FR). BALL, Maria
[VE/VE]; Avenue Cardenal Quintera, Res. Cardenal
Quintero, Edif. 10, Piso 4, Apto 42, Merida. Edo.,
Merida (VE). SEZONOV, Guennadi [RU/FR]; 16, rue
Saint Sauveur, F-75002 Paris (FR). TUPHILE, Karine
[FR/FR]; 39/41, boulevard Dubreuil, F-91400 Orsay (FR).
FROSTEGARD, Asa [NO/NO]; Flateby Skogsvei 7,
N-1450 Nesoddtangen (NO).

(74) Mandataire: BOUVET, Philippe; Aventis Pharma S.A.,
Direction Brevets, 20, avenue Raymond Aron, F-92165
Antony Cedex (FR).

(81) États désignés (national): AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE,
DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU,
ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO,
NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) États désignés (régional): brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), brevet eurasien
(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen

[Suite sur la page suivante]

(54) Title: METHOD FOR OBTAINING NUCLEIC ACIDS FROM AN ENVIRONMENT SAMPLE, RESULTING NUCLEIC
ACIDS AND USE IN SYNTHESIS OF NOVEL COMPOUNDS

(54) Titre: PROCEDE D'OBTENTION D'ACIDES NUCLEIQUES A PARTIR D'UN ECHANTILLON DE L'ENVIRONNE-
MENT, ACIDES NUCLEIQUES AINSI OBTENUS ET LEUR APPLICATION A LA SYNTHESE DE NOUVEAUX COMPOSES

(57) Abstract: The invention concerns a method for preparing nucleic acids from an environment sample, more particularly a method
for obtaining a library of nucleic acids from a sample. The invention also concerns nucleic acids of nucleic acid libraries obtained
by said method their use in the synthesis of novel compounds, in particular novel compounds of therapeutic interest. The invent
further concerns novel means used in the method for obtaining said nucleic acids, such as novel vectors and novel processes for
preparing such vectors or recombinant host cells containing said nucleic acid. Finally, the invention concerns methods for detecting
a nucleic acid of interest within a library of nucleic acids resulting from said method, and nucleic acids detected by said method and
polypeptides encoded by said nucleic acids.

(57) Abrégé: La présente invention concerne un procédé de préparation d'acides nucléiques à partir d'un échantillon de l'environ-
nement, plus particulièrement un procédé d'obtention d'une collection d'acides nucléiques à partir d'un échantillon. L'invention
est également relative aux acides nucléiques ou aux collections d'acides nucléiques obtenus selon le procédé et leur application à
la synthèse de nouveaux composés, notamment de nouveaux composés d'intérêt thérapeutique. L'invention a également pour objet
les moyens nouveaux mis en oeuvre dans le procédé d'obtention d'acides nucléiques ci-dessus, tels que de nouveaux vecteurs et des
nouveaux procédés de préparation de tels vecteurs ou encore des cellules hôtes recombinantes comprenant un acide nucléique de
l'invention. L'invention concerne encore des procédés pour détecter un acide nucléique d'intérêt au sein d'une collection d'acides
nucléiques obtenus selon le procédé ci-dessus, ainsi que les acides nucléiques détectés par un tel procédé et les polypeptides codés
par de tels acides nucléiques.

WO 01/40497 A2



(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

Publiée:

— *Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.*

Procédé d'obtention d'acides nucléiques à partir d'un échantillon de l'environnement, acides nucléiques ainsi obtenus et leur application à la synthèse de nouveaux composés

5 La présente invention concerne un procédé de préparation d'acides nucléiques à partir d'un échantillon de l'environnement, plus particulièrement un procédé d'obtention d'une collection d'acides nucléiques à partir d'un échantillon. L'invention est également relative aux acides nucléiques ou aux collections d'acides nucléiques obtenus
10 selon le procédé et leur application à la synthèse de nouveaux composés, notamment de nouveaux composés d'intérêt thérapeutique.

 L'invention a également pour objet les moyens nouveaux mis en oeuvre dans le procédé d'obtention d'acides nucléiques ci-dessus, tels que de nouveaux vecteurs et des nouveaux procédés de préparation
15 de tels vecteurs ou encore des cellules hôtes recombinantes comprenant un acide nucléique de l'invention.

 L'invention concerne encore des procédés pour détecter un acide nucléique d'intérêt au sein d'une collection d'acides nucléiques obtenus selon le procédé ci-dessus, ainsi que les acides nucléiques
20 détectés par un tel procédé et les polypeptides codés par de tels acides nucléiques.

 L'invention a également trait à des acides nucléiques obtenus et détectés selon les procédés ci-dessus, en particulier des acides nucléiques codant pour une enzyme participant à la voie de biosynthèse
25 d'antibiotiques tels que les β -lactames, les aminoglycosides, les nucléotides hétérocycliques ou encore des polykétides ainsi que l'enzyme codée par ces acides nucléiques, les polykétides produits grâce à l'expression de ces acides nucléiques et enfin des compositions pharmaceutiques comprenant une quantité pharmacologiquement active
30 d'un polykétide produit grâce à l'expression de tels acides nucléiques.

 Depuis la découverte de la production de la streptomycine par les actinomycètes, la recherche de nouveaux composés d'intérêt thérapeutique, et tout particulièrement de nouveaux antibiotiques, a eu recours de manière accrue à des méthodes de criblage des métabolites
35 produits par les micro-organismes du sol.

De telles méthodes consistent principalement à isoler les organismes de la microflore tellurique, à les cultiver sur des milieux nutritifs spécialement adaptés, puis à détecter une activité pharmacologique dans les produits retrouvés dans les surnageants de culture ou dans les lysats cellulaires ayant, le cas échéant, subi au préalable une ou plusieurs étapes de séparation et/ou de purification.

Ainsi, les méthodes d'isolement et de culture in vitro des organismes constituant la microflore tellurique ont permis, à la date d'aujourd'hui, de caractériser environ 40.000 molécules, dont environ la moitié présente une activité biologique.

Des produits majeurs ont été caractérisés selon de telles méthodes de culture in vitro, tels que des antibiotiques (pénicilline, érythromycine, actinomycine, tétracycline, céphalosporine), des anticancéreux, des anticholestérolémiants ou encore des pesticides.

Les produits d'intérêt thérapeutique d'origine microbienne connus à ce jour proviennent majoritairement (environ 70%) du groupe des actinomycètes et plus particulièrement du genre *Streptomyces*. Toutefois, d'autres composés thérapeutiques, tels que les teicoplanines, la gentamycine et les spinosines, ont été isolés à partir de micro-organismes de genres plus difficiles à cultiver tels que *Micromonospora*, *Actinomadura*, *Actinoplanes*, *Nocardia*, *Streptosporangium*, *Kitasatospora* ou encore *Saccharomonospora*.

Mais la pratique illustre le fait que la caractérisation de nouveaux produits naturels synthétisés par les organismes de la microflore du sol est restée limitée, en partie du fait que l'étape de culture in vitro aboutit le plus souvent à une sélection d'organismes déjà connus antérieurement.

Les méthodes de séparation et de culture in vitro des organismes telluriques en vue d'identifier de nouveaux composés d'intérêt présentent donc de nombreuses limites.

Chez les actinomycètes, par exemple, le taux de redécouverte d'antibiotiques déjà connus antérieurement est d'environ 99%. En effet, des techniques de microscopie en fluorescence ont permis de dénombrer plus de 10^{10} cellules bactériennes dans 1g de sol, alors que

seulement 0,1 à 1% de ces bactéries peuvent être isolées après ensemencement sur des milieux de culture.

A l'aide de techniques de cinétique de réassociation d'ADN, il a pu être montré qu'entre 12.000 et 18.000 espèces bactériennes peuvent
5 être contenues dans 1g de sol, alors qu'à ce jour, seuls 5000 micro-organismes non eucaryotes ont été décrits, tout habitat confondu.

Des études d'écologie moléculaire ont permis d'amplifier et cloner de nombreuses séquences nouvelles d'ADNr 16S à partir d'ADN de l'environnement.

10 Les résultats de ces études ont conduit à tripler le nombre de divisions bactériennes caractérisées antérieurement.

A la date d'aujourd'hui, les bactéries sont subdivisées en 40 divisions, certaines d'entre elles n'étant constituées que par des bactéries ne pouvant être cultivées. Ces derniers résultats témoignent de
15 l'ampleur de la biodiversité microbienne restée inexploitée à ce jour.

Des travaux récents ont tenté de surmonter les nombreux obstacles à l'accès à la biodiversité de la microflore du sol, dont notamment l'étape de culture in vitro préalable à l'isolement et la caractérisation de composés d'intérêt industriel, surtout d'intérêt
20 thérapeutique.

Des méthodes ont ainsi été mises au point qui incluent une étape d'extraction de l'ADN des organismes telluriques, le cas échéant après un isolement préalable des organismes contenus dans les échantillons de sol.

25 L'ADN ainsi extrait, après lyse des cellules bactériennes sans étape préalable de culture in vitro, est cloné dans des vecteurs utilisés pour transfecter des organismes hôtes, afin de constituer des banques d'ADN provenant de bactéries du sol.

Ces banques de clones recombinants sont utilisées pour
30 détecter la présence de gènes codant pour des composés d'intérêt thérapeutique ou alternativement pour détecter la production de composés d'intérêt thérapeutique par ces clones recombinants.

Toutefois, les méthodes d'accès direct à l'ADN de la microflore du sol, décrites dans l'état de la technique présentent des inconvénients
35 lors de la mise en oeuvre de chacune des étapes décrites ci-dessus, de

nature à affecter considérablement la quantité et la qualité du matériel génétique obtenu et exploitable.

L'état de la technique concernant chacune des étapes de construction de banques d'ADN provenant d'échantillons de sol est
5 détaillé ci-après, ainsi que les inconvénients techniques identifiés par le demandeur et qui ont été surmontés selon la présente invention.

1. Etape d'extraction de l'ADN à partir d'un échantillon du 10 **sol.**

1.1 Extraction directe d'ADN de l'environnement.

Il s'agit pour l'essentiel d'un procédé mettant en oeuvre des
15 techniques d'extraction d'ADN réalisées directement sur l'échantillon dans l'environnement, le plus souvent après une lyse in situ préalable des organismes de l'échantillon.

De telles techniques ont été mises en oeuvre sur des échantillons provenant de milieux aquatiques, que ce soit d'eau douce
20 ou marine. Elles comprennent une première étape de concentration préalable des cellules présentes librement ou sous forme de particules, consistant en général en une filtration de grands volumes d'eau sur différents dispositifs de filtration, par exemple filtration classique sur membrane, filtration tangentielle ou rotationnelle ou encore ultrafiltration.

25 La taille des pores est comprise entre 0,22 et 0,45 mm et nécessite souvent une préfiltration dans le but d'éviter des colmatages dus au traitement de grands volumes.

Dans un second temps, les cellules récoltées sont lysées directement sur les filtres dans des petits volumes de solutions, par
30 traitement enzymatique et/ou chimique.

Cette technique est par exemple illustrée par les travaux de STEIN et al. , 1996, Journal of Bacteriology, Vol.178 (3): 591-599 qui décrit le clonage de gènes codant pour de l'ADN ribosomal et pour un facteur d'élongation de la transcription (EF 2) à partir d'*Archaeobactéries*
35 du plancton marin.

Des techniques d'extraction directe d'ADN à partir d'échantillons de sol ou de sédiment ont été également décrites, basées sur des protocoles de lyse physique, chimique ou enzymatique réalisée in situ.

5 Par exemple, le brevet US N°5,824,485 (Chromaxome Corporation) décrit une lyse chimique des bactéries directement sur l'échantillon prélevé par addition d'un tampon de lyse chaud à base d'isothiocyanate de guanidium.

10 La demande Internationale n°WO 99/20.799 (WISCONSIN ALUMNI RESEARCH FOUNDATION) décrit une étape de lyse des bactéries in situ à l'aide d'un tampon d'extraction contenant une protéase et du SDS.

D'autres techniques ont également été utilisées telles que la réalisation de plusieurs cycles de congélation-décongélation de l'échantillon puis pressage de l'échantillon décongelé à haute pression. 15 Ont été également utilisées des techniques de lyse des bactéries à l'aide d'une succession d'étapes de sonication, de chauffage par micro-ondes et de chocs thermiques (PICARD et al. (1992).

Toutefois, les techniques d'extraction directe d'ADN de l'état de la technique décrites ci-dessus ont une efficacité très variable du point 20 de vue quantitatif et qualitatif.

Ainsi, les traitements chimiques ou enzymatiques in situ de l'échantillon ont le désavantage de ne lyser que certaines catégories de micro-organismes du fait de la résistance sélective des différents micro-organismes indigènes à l'étape de lyse en raison de leur morphologie 25 hétérogène.

Ainsi, les bactéries à Gram-positif résistent à un traitement à chaud au détergent SDS alors que la quasi-totalité des cellules à Gram-négatif sont lysées .

30 En outre, certains des protocoles d'extraction directe décrits ci-dessus favorisent l'adsorption des acides nucléiques extraits sur les particules minérales de l'échantillon, réduisant ainsi significativement la quantité d'ADN accessible.

Par ailleurs, si certains protocoles de l'état de la technique divulguent une étape de traitement mécanique pour lyser les micro- 35

organismes de l'échantillon prélevé, une telle étape de lyse mécanique est systématiquement effectuée en milieu liquide dans un tampon d'extraction, ce qui ne permet pas une bonne homogénéisation de l'échantillon de départ sous la forme de particules fines permettant une
5 accessibilité maximale à la diversité des organismes présents dans l'échantillon. Des essais de broyage ont également été effectués sur échantillon de sol brut à l'aide de billes de verre, mais la quantité d'ADN extrait était faible.

Il a été observé selon l'invention qu'une première étape de lyse
10 mécanique in situ en milieu liquide avait des effets négatifs sur la quantité d'ADN susceptible d'être extrait.

La quantité d'ADN directement utilisable pour le clonage dans des vecteurs recombinants est également tributaire des étapes de purification subséquentes à son extraction.

15 Dans l'état de la technique, l'ADN extrait est ensuite purifié, par exemple par l'utilisation de polyvinylpyrrolidone, par une précipitation en présence d'acétate d'ammonium ou de potassium, par des centrifugations sur gradient de chlorure de césium, ou encore des techniques chromatographiques, notamment sur support
20 d'hydroxyapatite, sur colonne échangeuse d'ions ou encore tamisage moléculaire ou par des techniques d'électrophorèse sur gel d'agarose.

Les techniques de purification d'ADN décrites antérieurement, surtout lorsque celles-ci sont combinées avec les techniques d'extraction d'ADN de l'environnement précitées, sont susceptibles de conduire à
25 une co-purification de l'ADN avec des composés inhibiteurs provenant de l'échantillon initial qui sont difficiles à éliminer.

La co-extraction de composés inhibiteurs avec l'ADN nécessite la multiplication du nombre d'étapes de purification ce qui conduit à des pertes importantes de l'ADN initialement extrait et réduit simultanément
30 la diversité du matériel génétique initialement contenu dans l'échantillon, ainsi que sa quantité.

Un autre but de l'invention a été de surmonter les inconvénients des protocoles de purification antérieurs et de mettre au point une étape de purification d'ADN permettant de maintenir de manière optimale la

diversité de l'ADN de l'échantillon initial, d'une part, et, de favoriser quantitativement son obtention, d'autre part.

Tout particulièrement, les améliorations qualitatives et quantitatives à la purification d'ADN sont maximales lorsqu'elles font appel à une combinaison d'un procédé d'extraction direct de l'ADN selon l'invention et d'un procédé de purification ultérieur, comme cela sera décrit ci-après.

1.2. Extraction indirecte d'ADN de l'environnement.

10

De telles techniques ont recours à une première étape de séparation des différents organismes de la microflore tellurique des autres constituants de l'échantillon de départ, préalablement à l'étape d'extraction de l'ADN proprement dite.

15

Dans l'état de la technique, la séparation préalable d'une fraction microbienne d'un échantillon de sol comprend le plus souvent une dispersion physique de l'échantillon par broyage de ce dernier en milieu liquide, par exemple en utilisant des dispositifs du type Waring Blender ou encore un mortier.

20

Il a également été décrit des dispersions chimiques, par exemple sur des résines échangeuses d'ions ou encore des dispersions à l'aide de détergents non spécifiques tels que le déoxycholate de sodium ou du polyéthylène glycol. Quel que soit le mode de dispersion, l'échantillon solide doit être mis en suspension dans de l'eau, du tampon phosphate ou une solution saline.

25

L'étape de dispersion physique ou chimique peut être suivie d'une centrifugation sur gradient de densité permettant la séparation des cellules contenues dans l'échantillon et des particules de ce dernier, étant entendu que les bactéries ont des densités inférieures à celles de la plupart des particules du sol.

30

L'étape de dispersion physique peut aussi être suivie alternativement d'une étape de centrifugation à faible vitesse ou encore une étape d'élutriation cellulaire.

35

L'ADN peut ensuite être extrait des cellules séparées par toutes les méthodes de lyse disponibles et être purifié par de nombreuses

méthodes, y compris les méthodes de purification décrites au paragraphe 1.1 précédent. Notamment, l'inclusion des cellules dans de l'agarose à bas point de fusion peut être réalisée afin de ménager la lyse.

5 Toutefois, les méthodes décrites dans l'état de la technique connues du demandeur ne donnent pas satisfaction du fait de la présence, dans les fractions contenant l'ADN extrait, de constituants indésirables de l'échantillon de départ ayant une influence significative sur la qualité et la quantité d'ADN final.

10 La présente invention se propose de résoudre les difficultés techniques rencontrées dans les procédés de l'art antérieur comme cela sera décrit ci-après.

2. Caractérisation moléculaire de l'ADN extrait.

15 Lorsque l'on désire construire une banque d'ADN à partir d'un échantillon de l'environnement, en particulier à partir d'un échantillon de sol, il est avantageux de vérifier la qualité et la diversité de la source d'ADN extrait et purifié préalablement à son insertion dans des vecteurs appropriés.

20 L'objectif d'une telle caractérisation moléculaire de l'ADN extrait et purifié est d'obtenir des profils représentant les proportions des différents taxons bactériens présents dans cet extrait d'ADN. La caractérisation moléculaire de l'ADN extrait et purifié permet de
25 déterminer si des artéfacts ont été introduits lors de la mise en oeuvre des différentes étapes d'extraction et de purification et, le cas échéant, si la diversité d'origine de l'ADN extrait et purifié est représentative de la diversité microbienne présente initialement dans l'échantillon, notamment dans l'échantillon de sol.

30 A la connaissance du demandeur, il est recouru dans l'état de la technique à des procédés d'hybridation quantitative mettant en oeuvre des sondes oligonucléotidiques spécifiques de différents groupes bactériens, appliqués directement à l'ADN extrait de l'environnement.

Malheureusement, une telle approche est peu sensible et ne permet pas de détecter des genres ou des groupes taxonomiques présents en faible abondance.

5 L'état de la technique décrit aussi des procédés de PCR quantitative, telle que la MPN-PCR ou encore la PCR quantitative par compétition. Toutefois, ces techniques présentent d'importants inconvénients.

Ainsi, la MPN-PCR est d'une utilisation complexe du fait de la multiplication des dilutions et des répétitions qui la rend inappropriée
10 pour un grand nombre d'échantillons ou de couples d'amorces.

Par ailleurs, la PCR quantitative par compétition est d'une mise en oeuvre difficile du fait de la nécessité de construire un compétiteur spécifique à l'ADN cible qui, en outre, n'induit pas de biais ou d'artéfacts dans la compétition proprement dite.

15 Il est ainsi proposé selon l'invention un procédé de précriblage d'une banque d'ADN provenant d'un échantillon de l'environnement qui est à la fois rapide, simple et fiable et permet de tester la qualité de l'ADN préalablement extrait et purifié et de déterminer ainsi l'intérêt de construire une banque de clones préparés à partir de cet ADN purifié de
20 départ.

3. Vecteurs pour le clonage de l'ADN extrait et purifié à partir d'un échantillon de l'environnement.

25 De nombreux vecteurs ont déjà été décrits dans l'état de la technique afin de cloner de l'ADN préalablement extrait d'un échantillon de l'environnement.

Ainsi, selon la description de la demande internationale n°WO 99/20.799, peuvent être utilisés des vecteurs viraux, des phages, des
30 plasmides, des phagemides, des cosmides, des phosmides, des vecteurs du type BAC (chromosome artificiel bactérien) ou encore le bactériophage P1, des vecteurs de type PAC (chromosome artificiel basé sur le bactériophage P1), des vecteurs du type YAC (chromosome artificiel de levure), des plasmides de levure ou tout autre vecteur

capable de maintenir et d'exprimer de manière stable un ADN génomique.

L'exemple 1 de la demande PCT n°WO 99/20.799 décrit la construction d'une banque d'ADN génomique par clonage dans un vecteur du type BAC.

A la connaissance du demandeur, aucune banque d'ADN provenant d'un échantillon de l'environnement n'avait encore été effectivement réalisée avec des vecteurs de type conjugatif, une telle technique étant rendue pour la première fois accessible et reproductible par l'homme du métier grâce à l'enseignement de la présente invention.

4. Hôtes cellulaires

Dans l'état de la technique, de nombreuses cellules hôtes ont été décrites comme pouvant être utilisées afin d'héberger les vecteurs contenant les inserts d'ADN provenant de l'ADN extrait et purifié à partir d'un échantillon de l'environnement.

Ainsi, la demande PCT N°WO 99/20.799 cite de nombreux hôtes cellulaires appropriés, tels que *Escherichia coli*, en particulier la souche DH 10B ou encore la souche 294 (ATCC 31446, la souche *E. coli* B, *E. Coli* X 1776 (ATCC N°31.537), *E.coli* DH5 α et *E.coli* W3110 (ATCC n°27.325).

Cette demande PCT cite également d'autres cellules hôtes appropriées telles que *Enterobacter*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Schigella* ou encore des souches du type bacillus telles que *B. subtilis* et *B. licheniformis* ainsi que les bactéries du genre *Pseudomonas*, *Streptomyces* ou *Actinomyces*.

Le brevet US N°5,824,485 cite en particulier la souche de *Streptomyces lividans* TK66 ou encore des cellules de levure telles que celles de *Saccharomyces pombe*.

5. Caractérisation de gènes d'intérêt dans des banques d'ADN provenant d'un échantillon de l'environnement.

La demande PCT N° WO 99/20.799 décrit une identification du phénotype de différents clones appartenant à la banque d'ADN de *B.cereus*, respectivement un clone produisant de l'hémolysine, un clone hydrolysant l'esculine ou encore un clone produisant un pigment orange.

5 Des techniques de mutagenèse basées sur l'utilisation d'un transposon codant pour l'enzyme pho A ont permis subséquemment d'isoler des clones mutés et de caractériser les séquences responsables des phénotypes observés.

L'article de STEIN et al. (1996) précité décrit l'utilisation
10 d'amorces spécifiques de l'ADN ribosomal afin d'amplifier l'ADN inséré dans les vecteurs hébergés par certains clones d'une banque d'ADN génomique d'Archaeobactéries de plancton marin et l'identification de plusieurs séquences codantes dans l'ADN ainsi amplifié.

L'article de BORSCHERT S. et al., (1992) décrit le criblage
15 d'une banque d'ADN génomique de *Bacillus subtilis* à l'aide de couples d'amorces hybridant avec des régions conservées de peptide synthétases connues afin d'identifier un ou plusieurs gènes correspondant dans le génome de *Bacillus subtilis*.

Cette technique a permis de détecter un fragment d'ADN
20 chromosomique d'environ 26 kb portant une partie de l'opéron de biosynthèse de la surfactine.

L'article de KAH-TONG S. et al.(1997) décrit le criblage d'une banque d'ADN provenant du sol à l'aide d'amorces hybridant avec des séquences conservées de l'opéron responsable de la voie de
25 biosynthèse des polykétides de type II et montre l'identification, au sein de cette banque d'ADN, de séquences apparentées au gène PKS- β . Cet article décrit aussi la construction de cassettes d'expression hybrides dans lesquelles la séquence de la sous-unité PKS- β , retrouvée naturellement dans l'opéron responsable de la biosynthèse des
30 polykétides, a été remplacée par différentes séquences apparentées retrouvées dans la banque d'ADN.

De même, l'article de HONG-FU et al. , (1995) décrit la construction de cassettes d'expression contenant les différentes phases de lecture ouverte de l'opéron responsable de la biosynthèse des
35 polykétides, les différentes cassettes d'expression ayant été construites

artificiellement en combinant les phases de lecture ouverte qui ne sont pas retrouvées ensemble naturellement dans le génome de *Streptomyces coelicolor*. Cet article montre que la combinaison, dans les cassettes d'expression artificielles, de cadres de lecture ouvert
5 originaux de différentes souches bactériennes permet la production de polykétides ayant différentes caractéristiques structurales et des activités antibiotiques plus ou moins grandes vis-à-vis de *Bacillus subtilis* et *Bacillus cereus*.

Les polykétides font partie d'une grande famille de produits
10 naturels de structure variable et possédant une grande diversité d'activités biologiques. Font partie des polykétides par exemple, les tétracyclines et l'érythromycine (antibiotiques), le FK506 (immunosuppresseur), la doxorubicine (agent anti-cancéreux), la monensine (un agent coccidiostatique) ainsi que l'ivermectine (un agent
15 antiparasitaire).

Ces molécules sont synthétisées grâce à des enzymes multifonctionnelles appelées polykétides synthases, qui catalysent des cycles de condensation répétés entre des acyl thioesters (en général des acétyl, propionyl, malonyl ou méthylmalonyl thioesters). Chaque cycle de
20 condensation aboutit à la formation, sur une chaîne croissante carbonée, d'un groupe β -céto qui peut ensuite subir, le cas échéant, une ou plusieurs séries d'étapes réductrices.

Compte-tenu de l'intérêt clinique important des polykétides, leur mécanisme commun de biosynthèse ainsi que le haut degré de
25 conservation observé entre les groupes de gènes codant pour les polykétides synthases, il s'est développé un intérêt accru pour le développement de nouveaux polykétides par génie génétique.

De nouveaux polykétides artificiels ont ainsi été produits par génie génétique, tels que la méderrhodine A ou la dihydrogranatirhodine.
30 La grande majorité des molécules nouvelles de polykétides obtenues par génie génétique sont très différentes, du point de vue structural, des polykétides correspondants naturels.

De l'état de la technique, il ressort ainsi qu'il existe un besoin d'obtention de nouveaux polykétides d'intérêt et tout particulièrement de
35 polykétides d'intérêt thérapeutique présentant notamment, par rapport à

leurs homologues naturels, un niveau accru d'activité antibiotique ou encore un spectre d'activité antibiotique différent, soit plus large que celui des polykétides connus, soit au contraire plus sélectif.

5 Ce besoin est, comme cela sera décrit ci-après, en partie comblé selon la présente invention.

DESCRIPTION DE L'INVENTION

10 L'invention concerne tout d'abord un procédé pour la construction de banques d'ADN provenant d'un échantillon de l'environnement, un tel échantillon pouvant être indifféremment un milieu aquatique (eau douce ou marine), un échantillon de sol (couche superficielle du sol, sous-sol ou sédiments), ou encore un échantillon d'organismes eucaryotes contenant une microflore associée, tel que par
15 exemple un échantillon provenant de plantes, d'insectes ou encore d'organismes marins et possédant une microflore associée.

La mise au point d'un procédé de construction d'une banque d'ADN d'un échantillon de l'environnement, et tout particulièrement d'un échantillon de sol, comprend des étapes critiques dont la mise en oeuvre
20 doit être nécessairement optimisée pour l'obtention d'une banque d'ADN dont le contenu en acides nucléiques d'intérêt répond aux objectifs initialement fixés.

Une première étape critique consiste en l'extraction et la purification ultérieure des acides nucléiques contenus initialement dans
25 l'échantillon, c'est-à-dire principalement des acides nucléiques contenus dans les divers organismes composant la microflore de cet échantillon.

La qualité de la purification de l'ADN extrait est déterminante sur le résultat obtenu.

Une seconde étape importante d'un procédé de construction
30 d'une banque d'acides nucléiques provenant d'un échantillon de l'environnement est l'évaluation de la diversité génétique des acides nucléiques extraits et purifiés. La mise au point d'une étape de réalisation simple et fiable de pré-criblage de l'ADN extrait et purifié afin de vérifier qu'il rend compte, au moins partiellement, de la diversité
35 phylogénétique des organismes présents initialement dans l'échantillon

de départ, permet en effet de déterminer l'intérêt ou non d'utiliser la source initiale d'ADN extrait et purifié pour la construction de la banque d'acides nucléiques proprement dite ou au contraire de ne pas poursuivre la construction de la banque d'acides nucléiques du fait d'artéfacts trop importants introduits au moment de l'extraction et de la purification des acides nucléiques. Il a en outre été identifié selon l'invention que la qualité des inserts introduits dans les vecteurs pour construire la banque est déterminante. Il a ainsi été déterminé que l'utilisation d'enzymes de restriction pour cliver l'ADN extrait et purifié à partir de l'échantillon de l'environnement était de nature à introduire des artéfacts ou "biais" dans la structure des inserts obtenus. En effet, l'ADN extrait du sol ou d'autres environnements, provenant en très grande majorité d'organismes non cultivables, est composé de molécules dont le taux de bases G et C est par définition inconnu et de plus variable en fonction de l'origine de ces organismes.

Une troisième étape critique est l'insertion des acides nucléiques extraits et purifiés dans des vecteurs capables d'intégrer des acides nucléiques de longueur choisie, d'une part, et, d'autre part, d'en permettre la transfection ou encore l'intégration dans le génome dans des hôtes cellulaires déterminés ainsi que, le cas échéant, d'en permettre l'expression dans de tels hôtes cellulaires.

Constituent des vecteurs d'intérêt, les vecteurs capables d'intégrer des acides nucléiques de grande taille, c'est-à-dire de taille supérieure à 100 kb lorsque l'objectif poursuivi consiste en un clonage et en une identification d'un opéron complet capable de diriger une voie complète de biosynthèse d'un composé d'intérêt industriel, en particulier d'un composé d'intérêt pharmaceutique ou agronomique.

DEFINITIONS

30

Au sens de la présente invention, on entend par "acides nucléiques", "polynucléotides" et "oligonucléotides" aussi bien des séquences d'ADN, d'ARN, que des séquences hybrides ARN/ADN de plus de 2 nucléotides, indifféremment sous la forme simple brin ou double brin.

35

Le terme " banque " ou " collection " est utilisé dans la présente description en référence indifféremment à un ensemble d'acides nucléiques extraits et, le cas échéant purifiés, provenant d'un échantillon de l'environnement, à un ensemble de vecteurs recombinants, chacun
5 des vecteurs recombinants de l'ensemble comprenant un acide nucléique provenant de l'ensemble d'acides nucléiques extraits et, le cas échéant purifiés précités, ainsi qu'à un ensemble de cellules hôtes recombinantes comprenant un ou plusieurs acides nucléiques provenant de l'ensemble des acides nucléiques extraits et, le cas échéant, purifiés
10 précités, lesdits acides nucléiques étant soient portés par un ou plusieurs vecteurs recombinants, soit intégrés dans le génome desdites cellules hôtes recombinantes.

On désigne par " échantillon de l'environnement " indifféremment un échantillon d'origine aquatique, par exemple d'eau
15 douce ou saline, ou un échantillon tellurique provenant de la couche superficielle d'un sol, de sédiments ou encore de couches inférieures du sol (sous-sol), ainsi que des échantillons d'organismes eucaryotes, le cas échéant multicellulaires, d'origine végétale, provenant d'organismes marins ou encore d'insectes et possédant une microflore associée, cette
20 microflore associée constituant des organismes d'intérêt.

On entend par " opéron " selon l'invention, un ensemble de cadres ouverts de lecture dont la transcription et/ou la traduction est co-réglée par un ensemble unique de signaux de régulation de la transcription et/ou de la traduction. Selon l'invention, un opéron peut
25 également comprendre lesdits signaux de régulation de la transcription et/ou de la traduction.

Par " voie métabolique " aux fins de l'invention ou encore " voie de biosynthèse " on entend un ensemble de réactions biochimiques anaboliques ou cataboliques réalisant la conversion d'une première
30 espèce chimique en une seconde espèce chimique.

Par exemple, une voie de biosynthèse d'un antibiotique est constituée de l'ensemble des réactions biochimiques convertissant des métabolites primaires en produits intermédiaires des antibiotiques, puis subséquentement en antibiotiques.

Par séquence de régulation placée "en phase" (en anglais operably linked) par rapport à une séquence nucléotidique dont l'expression est recherchée, on signifie que la ou les séquences de régulation de la transcription sont localisées, par rapport à la séquence
5 nucléotidique d'intérêt dont l'expression est recherchée, de manière à permettre l'expression de ladite séquence d'intérêt, la régulation de la dite expression étant dépendante de facteurs interagissant avec les séquences nucléotidiques régulatrices.

Selon une autre terminologie, on peut dire également que la
10 séquence nucléotidique d'intérêt dont l'expression est recherchée est placée "sous le contrôle" des séquences nucléotidiques régulatrices de la transcription.

Le terme "isolé" au sens de la présente invention désigne un matériel biologique qui a été soustrait à son environnement originel
15 (l'environnement dans lequel il est localisé naturellement).

Par exemple, un polynucléotide ou un polypeptide présent à l'état naturel dans un organisme (virus, bactérie, champignon, levure, plante ou animal) n'est pas isolé. Le même polypeptide séparé de son environnement naturel ou le même polynucléotide séparé des acides
20 nucléiques adjacents au sein desquels il est naturellement inséré dans le génome de l'organisme, est isolé.

Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel polynucléotide peut être inclus dans une composition et demeure néanmoins à l'état isolé, du fait que le vecteur ou la composition ne
25 constitue pas son environnement naturel.

Le terme "purifié" ne nécessite pas que le matériel soit présent sous une forme de pureté absolue, exclusif de la présence d'autres composés. Il s'agit plutôt d'une définition relative.

Un polypeptide ou un polynucléotide est à l'état purifié après
30 purification du matériel de départ d'au moins un ordre de grandeur, de préférence 2 ou 3 et préférentiellement 4 ou 5 ordres de grandeur.

Le "pourcentage d'identité" entre deux séquences de nucléotides ou d'acides aminés, au sens de la présente invention, peut être déterminé en comparant deux séquences alignées de manière
35 optimale, à travers une fenêtre de comparaison.

La partie de la séquence nucléotidique ou polypeptide dans la fenêtre de comparaison peut ainsi comprendre des additions ou des délétions (par exemple des "gaps") par rapport à la séquence de référence (qui ne comprend pas ces additions ou ces délétions) de manière à obtenir un alignement optimal des deux séquences.

Le pourcentage est calculé en déterminant le nombre de positions auquel une base nucléique ou un résidu d'acide aminé identique est observé pour les deux séquences (nucléique ou peptidique) comparées, puis en divisant le nombre de positions auquel il y a identité entre les deux bases ou résidus d'acides aminés par le nombre total de positions dans la fenêtre de comparaison, puis en multipliant le résultat par 100 afin d'obtenir le pourcentage d'identité de séquence.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus contenus dans le package de la Société WISCONSIN GENETICS SOFTWARE PACKAGE, GENETICS COMPUTER GROUP (GCG), 575 Science Doctor, Madison, WISCONSIN.

A titre d'illustration, le pourcentage d'identité de séquence pourra être effectué à l'aide du logiciel BLAST (versions BLAST 1.4.9 de Mars 1996, BLAST 2.0.4. de Février 1998 et BLAST 2.0.6. de Septembre 1998), en utilisant exclusivement les paramètres par défaut (S.F. Altschul et al., J. Mol. Biol. 1990 215: 403-410, S. F. Altschul et al., Nucleic Acids Res. 1997 25: 3389-3402). Blast recherche des séquences similaires/homologues à une séquence "requête" de référence, à l'aide de l'algorithme d'Altschul et al. La séquence requête et les bases de données utilisées peuvent être peptidiques ou nucléiques, toute combinaison étant possible.

EXTRACTION ET PURIFICATION D'ACIDES NUCLEIQUES PROVENANT D'UN ECHANTILLON DE L'ENVIRONNEMENT.

1. Extraction directe d'acides nucléiques

Il a été montré selon la présente invention que, pour l'obtention d'une banque d'acides nucléiques provenant d'organismes contenus

dans un échantillon du sol, il était important de créer des conditions dans lesquelles, d'une part, les différents organismes de l'échantillon sont rendus accessibles aux étapes ultérieures d'extraction des acides nucléiques et, d'autre part, que l'étape initiale de traitement de l'échantillon de sol permette une lyse mécanique maximale des organismes de l'échantillon de nature à rendre directement accessibles les acides nucléiques de ces organismes, principalement l'ADN génomique et plasmidique, aux tampons utilisés pour les étapes ultérieures d'extraction.

Il a été ainsi démontré selon l'invention qu'une accessibilité maximale des acides nucléiques provenant des micro-organismes d'un échantillon du sol était atteinte par un broyage poussé et à sec de l'échantillon de sol préalablement séché afin d'obtenir des micro-particules. Le demandeur a ainsi déterminé que le séchage de l'échantillon de sol préalable à tout traitement ultérieur provoque une diminution significative de la cohésion de l'échantillon de sol brut et favorise en conséquence sa désagrégation ultérieure sous la forme de micro-particules, lorsqu'un traitement par broyage approprié est opéré.

De manière surprenante, le demandeur a montré que des micro-particules d'échantillons de sol sec réunissaient des propriétés physico-chimiques favorables à l'extraction d'une quantité optimale d'acides nucléiques qui, dans leur nature, pouvaient être représentatifs de la diversité génétique des organismes présents initialement dans l'échantillon de sol de départ. Il a été montré en particulier que le procédé d'extraction directe d'acides nucléiques selon l'invention permettait l'extraction d'ADN provenant de micro-organismes rares, tels certains *Streptomyces* rares ou des micro-organismes sporulés.

Par "micro-particules" de l'échantillon de sol aux fins de la présente invention, on entend des particules dérivées de l'échantillon ayant une taille moyenne d'environ 50 μm , c'est à dire comprise en moyenne entre 45 et 55 μm .

Selon l'invention, les micro-particules sont obtenues à partir d'échantillons de sol préalablement séchés ou dessiqués puis broyés jusqu'à l'obtention de micro-particules de taille moyenne comprise entre

2µm et 50µm, avant remise en suspension dans un milieu tampon liquide des micro-particules obtenus.

Un tel milieu tampon liquide peut consister en un tampon d'extraction d'acides nucléiques, en particulier un tampon d'extraction d'ADN conventionnel bien connu de l'homme du métier.

Le broyage de l'échantillon de sol en micro-particules a pour double fonction de lyser mécaniquement la majorité des organismes présents dans l'échantillon de sol initial et de rendre accessibles les organismes non lysés par ce traitement mécanique à des étapes facultatives ultérieures de lyse chimique et/ou enzymatique.

Ainsi, un premier objet de l'invention consiste en un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes, ledit procédé comprenant une première étape (I-(a)) d'obtention de micro-particules par broyage de l'échantillon de sol préalablement séché ou dessiqué, puis mise en suspension des micro-particules dans un milieu tampon liquide.

De manière tout à fait préférée, l'étape de broyage est réalisée à l'aide d'un dispositif à billes d'agate ou de tungstène ou encore à l'aide d'un dispositif à anneaux de tungstène. Ces dispositifs sont préférés car la dureté de matériaux comme l'agate ou le tungstène facilite significativement l'obtention des micro-particules de la taille spécifiée ci-dessus. Pour cette raison, on ne choisira pas préférentiellement, voire on évitera, un recours à un dispositif de broyage à billes de verre, qui s'est révélé beaucoup moins efficace.

Le séchage ou la classification de l'échantillon de sol peut-être réalisée par toute méthode connue de l'homme du métier. Par exemple, l'échantillon de sol brut peut être séché à température ambiante pendant une durée de 24 à 48 heures.

Comme indiqué précédemment, le milieu tampon liquide peut consister en un milieu d'extraction de l'ADN présent dans les micro-particules. On utilisera de manière tout à fait préférée un tampon d'extraction désigné TENP contenant respectivement 50 mM tris, 20 mM EDTA, 100 mM NaCl et 1% (poids/volume) de polyvinylpyrrolidone, à pH 9,0.

Le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol est en outre caractérisé en ce que l'étape d'obtention de micro-particules par broyage de l'échantillon de sol préalablement séché ou dessiqué est suivie d'une étape I-(b) d'extraction des acides nucléiques présents dans les micro-particules.

Il est constant que l'extraction des acides nucléiques est accompagnée d'une co-extraction de composés et/ou de constituants du sol indésirables nécessitant la purification ultérieure des acides nucléiques extraits, une telle étape de purification ultérieure devant être à la fois suffisamment sélective pour permettre l'élimination des composés et/ou constituants du sol indésirables et d'un rendement suffisant pour entraîner une perte faible en quantité de l'ADN préalablement extrait.

Il a été montré selon l'invention qu'une étape de purification de l'ADN extrait des micro-particules de l'échantillon de sol répondant aux critères de sélectivité et de rendement définis ci-dessus, comprend un traitement de l'ADN extrait par une combinaison de deux étapes successives de chromatographie, respectivement une chromatographie sur tamis moléculaire et une chromatographie d'échange d'anions.

Selon une autre caractéristique du procédé ci-dessus, l'étape I-(b) d'extraction des acides nucléiques est suivie d'une étape I-(c) de purification des acides nucléiques extraits à l'aide des deux étapes de chromatographie suivantes:

- passage de la solution contenant les acides nucléiques sur un tamis moléculaire, puis récupération des fractions d'élution enrichies en acides nucléiques;

- passage des fractions d'élution enrichies en acides nucléiques sur un support de chromatographie d'échange d'anions, puis récupération des fractions d'élution contenant les acides nucléiques.

La nature et l'ordre des étapes de chromatographie ci-dessus sont essentiels à une bonne sélectivité et un excellent rendement de

l'étape de purification de l'ADN préalablement extrait des micro-particules de l'échantillon du sol préalablement séché ou dessiqué.

De manière très avantageuse, le support chromatographique du type " tamis moléculaire " de l'étape de purification d'acides nucléiques ci-dessus consiste en un support chromatographique de type Sephacryl® S400 HR ou un support chromatographique de caractéristiques équivalentes.

De manière tout à fait préférée, le support chromatographique d'échange d'anions utilisé lors de la seconde étape de purification de l'ADN extrait est un support de type Elutip® d, ou un support chromatographique de caractéristiques équivalentes.

En combinant les étapes I-(a) d'obtention de micro-particules de l'échantillon de sol sec, I-(b) d'extraction des acides nucléiques présents dans les micro-particules et I-(c) de purification par les étapes chromatographiques décrites ci-dessus, il a été possible selon l'invention d'extraire directement l'ADN du sol sans purification préalable des cellules des organismes contenus initialement dans l'échantillon, tout en évitant la co-extraction de contaminants du sol, tels que par exemple les acides humiques qui est observée avec les procédés de l'état de la technique.

Les contaminants, tels que les acides humiques affectent sévèrement les analyses et les utilisations subséquentes des acides nucléiques dont la purification est recherchée.

Selon le procédé ci-dessus, il est en outre possible d'accéder aux acides nucléiques contenus dans les organismes qui n'ont pas été lysés mécaniquement au cours de l'étape I-(a) d'obtention de micro-particules de l'échantillon de sol, dans le but d'obtenir une collection quasi-exhaustive de la diversité génétique des acides nucléiques présents initialement dans l'échantillon de sol. Ainsi, les micro-particules de l'échantillon de sol peuvent faire l'objet d'étapes ultérieures de traitement de lyse chimique, enzymatique ou physique, ou encore d'une combinaison de traitements chimiques, enzymatiques ou physiques.

Selon un premier aspect, le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol selon

l'invention, peut être en outre caractérisé en ce que l'étape I-(a) est suivie des étapes suivantes:

- traitement de la suspension de sol dans un milieu tampon
5 liquide par sonication;

- extraction et récupération des acides nucléiques.

De manière préférée, on aura recours, pour un traitement par
10 sonication, à un dispositif de type à micro-pointe en titane, tel que le dispositif 600 W Vibracell Ultrasonicator commercialisé par la Société Bioblock ou encore un sonicateur de type Cup Horn.

De manière tout à fait préférée, l'étape de sonication est
réalisée à une puissance de 15 W pendant une durée de 7 à 10 min et
15 comprend des cycles successifs de sonication, la sonication proprement dite étant réalisée pendant 50% de la durée de chaque cycle.

Selon un second aspect, le procédé ci-dessus peut être en
outre caractérisé en ce que l'étape I-(a) est suivie des étapes suivantes:

- traitement de la suspension de sol dans un milieu tampon
20 liquide par sonication;

- incubation de la suspension à 37°C après sonication en
présence de lysozyme et d'achromopeptidase;

- addition de SDS avant centrifugation et précipitation des
25 acides nucléiques;

- récupération des acides nucléiques précipités.

30

De préférence, l'étape d'incubation en présence de lysozyme et
d'achromopeptidase sera réalisée à une concentration finale de 0,3
mg/ml de chacune des deux enzymes, préférentiellement pendant 30
minutes à 37°C.

De manière préférée, le SDS sera utilisé à une concentration finale de 1% et pendant un temps d'incubation de 1 heure à la température de 60°C avant centrifugation et précipitation.

Selon un troisième aspect, le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol ci-dessus est en outre caractérisé en ce que l'étape I-(a) est suivie des étapes suivantes:

- homogénéisation de la suspension de sol avec une étape de mixage violent (vortex) suivie d'une étape de simple agitation;
- congélation de la suspension homogène suivie d'une décongélation ;
- traitement par sonication de la suspension après décongélation;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- addition de SDS avant centrifugation et précipitation des acides nucléiques;
- récupération des acides nucléiques.

De manière préférée, les suspensions de micro-particules de sol sont passées au vortex puis homogénéisées par une agitation douce sur un agitateur à rotation circulaire pendant une durée de deux heures avant d'être congelées à -20°C.

Préférentiellement, les suspensions sont à nouveau agitées violemment par vortex pendant 10 minutes, après décongélation et avant l'étape de sonication.

Il va sans dire que les acides nucléiques extraits par les modes de réalisation du procédé d'extraction directe d'acides nucléiques décrit ci-dessus sont préférentiellement purifiés selon l'étape de purification constituée d'un premier passage sur tamis moléculaire puis un passage subséquent des fractions d'élution obtenues à l'issue de la chromatographie sur tamis moléculaire sur un support chromatographique d'échange d'anions.

2. Extraction indirecte des acides nucléiques

Selon un second mode de réalisation du procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement, selon l'invention, ledit échantillon de l'environnement subit un premier traitement de nature à permettre la séparation des organismes, contenus dans cet échantillon, des autres macro-constituants de l'échantillon.

Ce second mode de réalisation du procédé de préparation d'une collection d'acides nucléiques selon l'invention favorise l'obtention d'acides nucléiques de grande taille, qui sont pratiquement impossibles à obtenir selon le premier mode de réalisation du procédé selon l'invention décrit ci-dessus, l'étape de lyse mécanique opérée pour l'obtention des micro-particules ayant également pour effet de casser physiquement les acides nucléiques de l'échantillon de sol ou des acides nucléiques contenus dans les organismes de l'échantillon de sol.

L'obtention d'acides nucléiques de grande taille a été recherchée par le demandeur dans le but d'isoler et de caractériser les acides nucléiques comprenant, au moins partiellement, l'ensemble des séquences codantes appartenant à un même opéron capable de diriger la biosynthèse d'un composé d'intérêt industriel.

De manière préférée, on obtient, en mettant en oeuvre le second mode de réalisation du procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol selon l'invention, des acides nucléiques ayant une taille supérieure à 100 kb, de préférence supérieure à 200, 250 ou 300 kb, et de manière tout à fait préférée d'acides nucléiques d'une taille supérieure à 400, 500 ou encore 600 kb.

Ce second mode de réalisation d'un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement selon l'invention est constitué d'une combinaison de quatre étapes successives destinées à l'obtention des acides nucléiques ayant les caractéristiques décrites ci-dessus.

Lorsque l'échantillon de l'environnement est un échantillon de sol, il a été montré selon l'invention qu'une première étape d'obtention d'une suspension par dispersion de l'échantillon de sol en milieu liquide

favorisait l'accessibilité des organismes contenus dans l'échantillon sans provoquer de lyse mécanique significative des cellules.

La première étape d'obtention d'une dispersion de l'échantillon de sol ci-dessus rend accessibles les organismes de l'échantillon au milieu extérieur et permet également une dissociation partielle des organismes de l'échantillon et des macro-constituants. Elle rend ainsi possible une séparation ultérieure des organismes contenus initialement dans l'échantillon des autres constituants de ce dernier.

Lorsque l'échantillon de l'environnement provient par exemple de végétaux, d'organismes marins ou d'insectes, un traitement préalable par broyage est nécessaire afin de rendre les organismes de la microflore associée accessible aux étapes ultérieures du procédé.

Ainsi, le présent procédé comprend une étape de séparation des organismes des autres constituants minéraux et/ou organiques obtenus précédemment par une centrifugation sur un gradient de densité. Les organismes ainsi séparés sont ensuite soumis à une étape de lyse puis d'extraction des acides nucléiques .

L'étape de centrifugation sur un gradient de densité a, de manière surprenante, permis de séparer les cellules d'organismes des particules de sol contenues dans la suspension de l'échantillon. On aurait en effet pu s'attendre à ce qu'une proportion des cellules soient entraînées avec les macro-particules au sein de la phase de gradient. En outre, il n'avait jamais été démontré jusqu'à présent qu'une centrifugation sur gradient de densité d'un échantillon de sol permettait de retrouver, à l'interface phase aqueuse/gradient, une population d'organismes représentative de la diversité des organismes présents dans l'échantillon de départ, du fait que ces organismes sont de volume, densité et forme extrêmement variables. On pouvait raisonnablement supposer qu'ils seraient retrouvés indifféremment au sein de la phase aqueuse, à l'interface phase aqueuse/gradient de densité et également au sein du gradient de densité lui-même.

Ainsi, l'homme du métier pouvait s'attendre à ce que des organismes présentant des densités plus faibles ou plus grandes que la densité du gradient de densité utilisé (densité du gradient de densité

comprise entre 1,2 et 1,5 g/ml , préférentiellement 1,3 g/ml) ne pouvait être récupérés, ce qui aurait eu pour effet d'introduire un biais dans la représentativité des organismes effectivement séparés et, par voie de conséquence, également dans la diversité des acides nucléiques extraits.

En outre, dans un mode de réalisation particulier du procédé, une étape de germination des spores, en particulier d'actinomycètes, est réalisée, ce qui a pour effet d'accroître de manière significative la quantité d'ADN d'actinomycètes récupérée.

La dernière étape consiste en une étape de purification des acides nucléiques ainsi extraits sur un gradient de chlorure de césium.

De manière surprenante, la purification des acides nucléiques sur le gradient de chlorure de césium permet une élimination substantielle, voire complète, des substances composant le gradient de densité. Cette caractéristique est déterminante en ce qui concerne l'utilisation ultérieure des acides nucléiques purifiés car le gradient de densité est connu comme un puissant inhibiteur enzymatique, capable le cas échéant d'inhiber l'activité catalytique des enzymes utilisées pour préparer l'insertion des acides nucléiques extraits dans des vecteurs.

Selon ce second mode de réalisation, le procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement contenant des organismes selon l'invention comprend la succession d'étapes suivantes:

(i) obtention d'une suspension par dispersion de l'échantillon de l'environnement en milieu liquide puis homogénéisation de la suspension obtenue par agitation douce;

(ii) séparation des organismes des autres constituants minéraux et/ou organiques de la suspension homogène obtenue à l'étape (i) par centrifugation sur un gradient de densité;

(iii) lyse des microorganismes séparés à l'étape (ii) et extraction des acides nucléiques ;

(iv) purification des acides nucléiques sur un gradient de chlorure de césium .

Préférentiellement, la suspension de l'échantillon de sol est
5 obtenue par dispersion de cet échantillon par broyage à l'aide d'un
dispositif de type Waring Blender ou un dispositif de caractéristiques
équivalentes. De manière tout à fait préférée, la suspension d'échantillon
est obtenue après trois broyages successifs d'une durée d'une minute
chacun dans un dispositif de type Waring Blender. De préférence,
10 l'échantillon broyé sera refroidi dans la glace entre chacun des broyages.

De manière préférée, les organismes sont ensuite séparés des
particules du sol par centrifugation sur un coussin de densité du type
" Nycodenz ", commercialisé par la Société Nycomed Pharma AS. (Oslo
, Norvège). Les conditions préférées de centrifugation sont de 10.000g
15 pendant 40 minutes à 4°C, avantageusement dans un rotor à godets
mobiles du type " rotor TST 28.38 " commercialisé par la Société
KONTRON.

L'anneau d'organismes localisé, après centrifugation, à
l'interphase de la phase supérieure aqueuse et de la phase inférieure de
20 Nycodenz est alors prélevé et lavé par centrifugation avant reprise du
culot cellulaire dans un tampon approprié.

L'étape (iii) de lyse des organismes séparés à l'étape (ii) décrite
ci-dessus peut être réalisée de toute manière connue de l'homme du
métier.

25 Avantageusement, les cellules sont lysées dans une solution
Tris 10 mM-EDTA 100mM à pH 8.0 en présence de lysozyme et
d'achromopeptidase, avantageusement pendant une heure à 37°C.

L'extraction proprement dite de l'ADN peut être
avantageusement réalisée par addition d'une solution de lauryl sarcosyl
30 (1% du poids final de la solution) en présence de protéinase K et
incubation de la solution finale à 37°C pendant 30 minutes.

Les acides nucléiques extraits à l'étape (iii) sont ensuite purifiés
sur un gradient de chlorure de césium. Préférentiellement, l'étape de
purification des acides nucléiques sur un gradient de chlorure de césium

est réalisée par centrifugation à 35.000 tours/minute pendant 36 heures, par exemple sur un rotor du type Kontron 65.13.

Selon un aspect particulier du procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention, lesdits acides nucléiques sont constitués majoritairement, sinon exclusivement, de molécules d'ADN.

Selon un autre aspect, les acides nucléiques peuvent être récupérés après inclusion des organismes, séparés sur un gradient de densité, dans un bloc d'agarose et lyse, par exemple chimique et/ou enzymatique, des organismes inclus dans le bloc d'agarose.

Un autre objet de l'invention consiste en une collection d'acides nucléiques constitués des acides nucléiques obtenus à l'étape II-(iv) du procédé de préparation d'une collection d'acides nucléiques selon l'invention ou encore obtenue à l'étape (c) ou une étape ultérieure du procédé de préparation d'une collection d'acides nucléiques selon l'invention.

L'invention est encore relative à un acide nucléique caractérisé en ce qu'il est contenu dans une collection d'acides nucléiques telle que définie ci-dessus.

Selon un premier aspect, un tel acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention est caractérisé en ce qu'il comprend une séquence nucléotidique codant au moins un opéron, ou une partie d'un opéron.

De manière tout à fait préférée, un tel opéron code pour la totalité ou une partie d'une voie métabolique.

L'exemple 9 décrit la construction d'une banque d'ADN génomique à partir d'une souche de *Streptomyces alboniger* et son clonage respectivement dans les cosmides navettes pOS700I et pOS700R. Il a été montré selon l'invention que dans la banque d'ADN réalisée dans le vecteur intégratif pOS700I neuf clones contiennent des séquences nucléotidiques appartenant à l'opéron responsable de la voie de biosynthèse de la puromycine. De même, il a pu être identifié au sein de la banque d'ADN réalisée dans le vecteur réplcatif pOS 700R douze

clones contenant des séquences nucléotidiques de l'opéron responsable de la voie de biosynthèse de la puromycine.

En particulier, certains cosmides intégratifs et replicatifs des banques réalisées présentent, après digestion par les endonucléases de restriction ClaI et EcoRV, un fragment d'une taille de 12 kb susceptible
5 de contenir la totalité des séquences de l'opéron responsable de la voie de biosynthèse de la puromycine.

Ainsi, selon un autre aspect, un acide nucléique selon l'invention contient, au moins en partie, des séquences nucléotidiques de
10 l'opéron responsable de la voie de biosynthèse de la puromycine.

L'exemple 2 ci-après décrit la construction d'une banque d'ADN selon un procédé conforme à la présente invention dans un vecteur pBluescript SK⁻ à partir d'un sol contaminé par du lindane.

Les vecteurs recombinants ont été transfectés dans des
15 cellules d'*Escherichia coli* DH10B puis les cellules transformées ont été cultivées dans un milieu de culture approprié en présence de lindane. Un criblage des clones de cellules transformées de la banque a permis de montrer que, sur 10.000 clones criblés, 35 d'entre eux présentaient un phénotype de dégradation du lindane. La présence du gène linA chez
20 ces clones a pu être confirmée par amplification PCR grâce à des amorces spécifiques de ce gène.

Ainsi, selon un autre aspect, l'invention concerne également un acide nucléique contenant une séquence nucléotidique de la voie métabolique provoquant la biodégradation du lindane.

25 Il est donc clairement démontré, comme décrit plus haut, qu'un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention ainsi qu'un procédé de préparation d'une collection de vecteurs recombinants contenant les acides nucléiques constitutifs de la collection d'acides
30 nucléiques précités était tout à fait apte à l'isolement et à la caractérisation de séquences nucléotidiques incluses dans un opéron.

Une démonstration supplémentaire de l'aptitude d'un procédé selon l'invention à l'identification de séquences nucléotidiques codantes impliquées dans une voie de biosynthèse régulée sous la forme d'un
35 opéron est en outre décrite plus loin: il s'agit du clonage et de la

caractérisation de séquences codant pour des polykétides synthases impliquées dans la voie de biosynthèse des polykétides, qui appartiennent à une famille de molécules dont certains représentants sont d'un intérêt thérapeutique majeur, en particulier antibiotique.

5 La présente invention a donc en outre pour objet un acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention, caractérisé en ce qu'il comprend la totalité d'une séquence nucléotidique codant pour un polypeptide.

 Selon un premier aspect, un acide nucléique constitutif d'une
10 collection d'acides nucléiques selon l'invention est d'origine procaryote.

 Selon un second aspect, un acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention provient d'une bactérie ou d'un virus.

 Selon un troisième aspect, un acide nucléique constitutif d'une
15 collection d'acides nucléiques selon l'invention est d'origine eucaryote.

 En particulier, un tel acide nucléique est caractérisé en ce qu'il provient d'un champignon, d'une levure, d'une plante ou d'un animal.

CARACTERISATION MOLECULAIRE DE LA COLLECTION D'ACIDES 20 **NUCLEIQUES EXTRAITS DU SOL.**

 Afin de surmonter les nombreux inconvénients techniques des méthodes de caractérisation des banques d'ADN extraits et purifiés à partir d'un échantillon de l'environnement qui ont été décrits dans la
25 partie de la description relative à l'état de la technique, le demandeur a mis au point un procédé simple et fiable permettant de caractériser qualitativement et semi-quantitativement les acides nucléiques obtenus à l'issue du procédé décrit ci-dessus.

 Le procédé selon l'invention consiste ainsi à amplifier
30 universellement un fragment de 700 pb localisé à l'intérieur d'une séquence d'ADN ribosomal de type 16 S, puis d'hybrider l'ADN amplifié avec une sonde oligonucléotidique de spécificité variable et enfin de comparer l'intensité d'hybridation de l'échantillon par rapport à une gamme étalon externe d'ADN de séquence ou d'origine connue.

L'amplification préalable à l'hybridation avec la sonde oligonucléotidique permet de quantifier des genres ou des espèces de micro-organismes peu abondants. De plus, l'amplification par des amorces universelles permet, lors de l'hybridation, d'utiliser une large
5 série de sondes oligonucléotidiques.

Ainsi, l'invention a en outre pour objet un procédé de détermination de la diversité des acides nucléiques contenus dans une collection d'acides nucléiques, et tout particulièrement d'une collection d'acides nucléiques provenant d'un échantillon de l'environnement,
10 préférentiellement d'un échantillon du sol, ledit procédé comprenant les étapes suivantes:

- mise en contact des acides nucléiques de la collection d'acides nucléiques à tester avec un couple d'amorces oligonucléotidiques hybridant à toute séquence d'ADN ribosomal 16 S
15 bactérien;

- réalisation d'au moins trois cycles d'amplification ;
- détection des acides nucléiques amplifiés à l'aide d'une sonde oligonucléotidique ou d'une pluralité de sondes oligonucléotidiques,
20 chaque sonde hybridant spécifiquement avec une séquence d'ADN ribosomal 16 S commune à un règne, un ordre, une sous-classe ou un genre bactérien;

- le cas échéant, comparaison des résultats de l'étape de détection précédente avec les résultats de détection, à l'aide de la sonde
25 ou de la pluralité de sondes d'acides nucléiques de séquence connue constituant une gamme étalon.

De manière préférée, un premier couple d'amorces hybridant avec des régions universellement conservées du gène de l'ARN ribosomal 16 S est constitué respectivement des amorces FGPS 612
30 (SEQ ID N°12) et FGPS 669 (SEQ ID N°13).

Un second mode de réalisation d'un couple d'amorces préféré selon l'invention est constitué du couple d'amorces universelles 63 f (SEQ ID N°22) et 1387r (SEQ ID N°23).

Selon un mode particulier de réalisation d'un procédé de
35 détermination de la diversité des acides nucléiques d'une collection

d'acides nucléiques, l'étape d'amplification à l'aide d'un couple d'amorces universelles peut être réalisée sur une collection de vecteurs recombinants dans chacun desquels a été inséré un acide nucléique de la collection d'acides nucléiques considérée, préalablement à l'étape d'hybridation avec les sondes oligonucléotidiques spécifiques d'un

5 règne, d'un ordre, d'une sous-classe ou d'un genre bactérien particulier.

Un tel procédé de détermination de la diversité des acides nucléiques contenus dans une collection est tout particulièrement applicable aux collections d'acides nucléiques obtenus conformément à

10 l'enseignement de la présente description.

Ainsi, l'exemple 3 détaille un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes comprenant une étape d'extraction indirecte d'ADN par dispersion d'un échantillon du sol préalablement à la séparation des

15 cellules sur gradient de Nycodenz, lyse des cellules puis purification de l'ADN sur gradient de chlorure de césium.

La collection d'acides nucléiques ainsi obtenue a été utilisée telle quelle ou sous la forme d'inserts dans des vecteurs de type cosmide dans un procédé d'amplification à l'aide des amorces

20 universelles de l'ADNr 16 S précitées, puis les ADN amplifiés ont été soumis à une étape de détection à l'aide de sondes oligonucléotidiques de séquences SEQ ID N°14 à SEQ ID N°21 qui sont présentées dans le tableau 4.

Les résultats montrent qu'un procédé de préparation d'une

25 collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention permet d'accéder à l'ADN de plus de 14% de la microflore tellurique totale, soit 2×10^8 cellules par gramme de sol, alors que la microflore totale cultivable ne représente qu'à peine 2% de la population microbienne totale.

30 Afin de déterminer la diversité phylogénétique d'une collection d'acides nucléiques préparés conformément à l'invention, 47 séquences du gène ARNr 16S ont été isolées et séquencées. Ces séquences correspondent respectivement aux séquences nucléotidiques SEQ ID N°60 à SEQ ID N°106.

Les acides nucléiques comprenant les séquences SEQ ID N° 60 à SEQ ID N° 106 font également partie de l'invention, ainsi que les acides nucléiques possédant au moins 99 %, préférentiellement 99,5% ou 99,8% d'identité en acides nucléiques avec les acides nucléiques comprenant les séquences SEQ ID N° 60 à SEQ ID N° 106. De telles séquences peuvent être utilisées notamment en tant que sondes pour cribler des clones d'une banque d'ADN et identifier ainsi ceux, parmi les clones de la banque, qui contiennent de telles séquences, ces séquences étant susceptibles d'être à proximité de séquences codantes d'intérêt, telles que des séquences codant pour des enzymes impliquées dans la voie de biosynthèse de métabolites antibiotiques, par exemple des polykétides.

La comparaison des séquences d'ARNr 16S à partir d'une banque d'ADN réalisée conformément à l'invention avec les séquences répertoriées dans la base données RDP (Maidak B.L., Cole J.R., Parker C.T., Garrity G.M., Larsen N., Li B., Lilburn T.G., McCaughey, M.J., Olsen G.J., Overbeek R., Pramanik S., Schmidt T.M., Tiedje J.M., Woese C.R. (1999) "A new projet of the RDP (Ribosomal Database Project)" Nucleic Acids Research Vol. 27: 171-173) ont permis de déterminer que les acides nucléiques contenus dans une collection d'acides nucléiques selon l'invention proviennent d' α -protéobactéries, de β -protéobactéries, de δ -protéobactéries, de γ -protéobactéries, d'actinomycètes ainsi que d'un genre apparenté à acidobactérium. Ces résultats, présentés dans le tableau 7 ainsi que par l'arbre phylogénétique de la figure 7 rendent compte de la grande diversité phylogénétique des acides nucléiques contenus dans une banque d'ADN préparée conformément au procédé selon l'invention.

VECTEURS DE CLONAGE ET/OU D'EXPRESSION

30

Chacun des acides nucléiques contenus dans une collection d'acides nucléiques préparés conformément à l'invention peut être inséré dans un vecteur de clonage et/ou d'expression.

A cette fin, tous types de vecteurs connus de l'état de la technique peuvent être utilisés, tels que des vecteurs viraux, des

phages, des plasmides, des phagemides, des cosmides, des phosmides, des vecteurs de type BAC, des bactériophages P1, des vecteurs de type BAC, des vecteurs de type YAC, des plasmides de levure ou encore tout autre vecteur connu de l'état de la technique par l'homme du métier.

On aura avantageusement recours selon l'invention à des vecteurs permettant une expression stable des acides nucléiques d'une banque d'ADN. A cette fin, de tels vecteurs incluent préférentiellement des séquences de régulation de la transcription qui sont localisées en phase ("operably linked") avec l'insert génomique de manière à permettre l'initiation et/ou la régulation de l'expression d'au moins une partie dudit insert d'ADN.

Il résulte de ce qui précède, que l'invention concerne encore un procédé de préparation d'une collection de vecteurs recombinants caractérisé en ce que les acides nucléiques obtenus à l'étape II-(iv) ou à l'étape I-(c) ou toute autre étape ultérieure d'un procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes selon l'invention sont insérés dans un vecteur de clonage et/ou d'expression.

Préalablement à leur insertion dans un vecteur de clonage et/ou d'expression, les acides nucléiques constitutifs d'une collection d'acides nucléiques selon l'invention peuvent être séparés en fonction de leur taille, par exemple par électrophorèse sur un gel d'agarose, le cas échéant après digestion à l'aide d'une endonucléase de restriction.

Selon un autre aspect, la taille moyenne des acides nucléiques constitutifs d'une collection d'acides nucléiques selon l'invention peut être rendue d'une taille sensiblement uniforme par la mise en oeuvre d'une étape de rupture physique préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

Une telle étape de rupture physique ou mécanique des acides nucléiques peut consister en des passages successifs de ces derniers, en solution, dans un canal métallique d'environ 0,4 mm de diamètre, par exemple le canal d'une aiguille de seringue ayant un tel diamètre.

La taille moyenne des acides nucléiques peut dans ce cas être comprise entre 30 et 40 kb de longueur.

La construction des vecteurs préférés selon l'invention est schématisée dans les figures 25 (cosmide intégratif conjugatif) et 26 (BAC intégratif).

Des vecteurs de clonage et/ou d'expression pouvant être
5 avantageusement utilisés aux fins d'insertion des acides nucléiques
contenus dans une collection ou banque d'ADN selon l'invention sont
notamment les vecteurs décrits dans le brevet européen N°EP-0 350
341 et dans le brevet US N°5 688 689, de tels vecteurs étant
spécialement adaptés à la transformation de souches d'actinomycètes.
10 De tels vecteurs contiennent, outre une séquence d'ADN de l'insert, une
séquence d'attachement att ainsi qu'une séquence d'ADN codant pour
une intégrase (séquence int) fonctionnelle dans les souches
d'actinomycètes.

Toutefois, il a été observé selon l'invention que certains
15 vecteurs de clonage et/ou d'expression présentaient des inconvénients
et que leur capacité fonctionnelle théorique n'était pas atteinte dans la
pratique.

Ainsi, il est apparu que le système d'intégration contenu dans
des vecteurs de l'état de la technique, et notamment dans les vecteurs
20 décrits dans le brevet européen n°EP 0 350 41 ne permettait pas en
réalité une bonne intégration de l'insert d'ADN de la banque au sein du
chromosome bactérien.

Partant de l'hypothèse que les déficits fonctionnels d'intégration
25 de tels vecteurs au sein du chromosome bactérien étaient dus à un
défaut dans l'expression du gène de l'intégrase présent dans ces
vecteurs, le demandeur a tout d'abord cherché à augmenter l'expression
du gène de l'intégrase en substituant au promoteur de la transcription
initial un promoteur de la transcription susceptible d'augmenter
30 significativement le nombre de transcrits de l'intégrase.

Les résultats ont été décevants et la fonction d'intégration au
chromosome de ces vecteurs n'a pas été améliorée.

De manière surprenante, il a été montré selon l'invention que
les difficultés d'expression de l'intégrase contenues dans cette famille de

vecteurs intégratifs ne se situait pas au niveau de la quantité d'expression des transcrits, mais au niveau de leur stabilité.

Selon une seconde hypothèse, le demandeur a pu montrer que le défaut de stabilité des transcrits de l'intégrase était causé par des
5 déficits dans la terminaison de la transcription de l'ARN messenger correspondant.

Le demandeur a alors inséré un site terminateur placé en aval de la séquence codant pour l'intégrase du vecteur de manière à obtenir un ARN messenger de taille déterminée. L'insertion d'un signal de
10 terminaison additionnel en aval de la séquence nucléotidique codant pour l'intégrase du vecteur a permis l'obtention d'une famille de vecteurs intégratifs de type cosmide et de type BAC .

Préférentiellement, le site terminateur est placé en aval du site d'attachement att.

15

En outre, le demandeur a mis au point de nouveaux vecteurs conjugatifs et de nouveaux vecteurs réplicatifs du type cosmide et de nouveaux vecteurs conjugatifs de type BAC qui peuvent
20 avantageusement être utilisés pour l'insertion des acides nucléiques constitutifs d'une collection d'acides nucléiques préparés selon le procédé de l'invention.

Lorsque l'insertion de fragments d'ADN de taille moyenne est recherchée, on utilise préférentiellement des vecteurs du type cosmide, capables de recevoir des inserts ayant une taille maximale d'environ 50
25 kb.

De tels vecteurs cosmidiques sont tout particulièrement adaptés pour l'insertion d'acides nucléiques constitutifs d'une collection d'acides nucléiques obtenus selon le procédé de l'invention comprenant une première étape d'extraction directe d'ADN par lyse mécanique des
30 organismes contenus dans l'échantillon de sol initial.

Lorsque l'insertion d'acides nucléiques de grande taille, en particulier d'acides nucléiques d'une taille supérieure à 100 kb, voire supérieure à 200, 300, 400, 500 ou 600 kb est recherchée, on aura alors recours préférentiellement à des vecteurs du type BAC capables de
35 recevoir des inserts d'ADN d'une telle taille.

De tels vecteurs de type BAC sont tout particulièrement adaptés pour l'insertion des acides nucléiques constitutifs d'une collection d'acides nucléiques obtenus conformément au procédé selon l'invention dans lequel la première étape est constituée d'une extraction
5 indirecte de l'ADN par séparation préalable des organismes contenus dans l'échantillon de sol initial et élimination des macro-constituants dudit échantillon de sol.

En particulier, des vecteurs du type BAC sont avantageusement mis en oeuvre pour l'insertion d'acides nucléiques de grande taille
10 contenant, au moins partiellement, la séquence nucléotidique d'un opéron.

Ainsi, le procédé de préparation d'une collection de vecteurs recombinants de clonage et/ou d'expression selon l'invention est en outre caractérisé en ce que le vecteur de clonage et/ou d'expression est
15 du type plasmide.

Selon un autre aspect, un tel procédé est caractérisé en ce que le vecteur de clonage et/ou d'expression est du type cosmide.

Selon un premier aspect, il peut s'agir d'un cosmide répliatif chez *E.coli* et intégratif chez *Streptomyces*. Un vecteur cosmidique tout à
20 fait préféré répondant à une telle définition est le cosmide pOS7001 décrit à l'exemple 3.

Selon encore un autre aspect, le vecteur cosmidique est conjugatif et intégratif chez *Streptomyces*.

De manière générale, des vecteurs conjugatifs de type cosmide
25 ou de type BAC, qui comprennent dans leurs séquences nucléotidiques un motif reconnu par la machinerie enzymatique cellulaire appelé "origine de conjugaison" sont utilisés chaque fois que l'on veut éviter un recours à des techniques de transformation lourdes et peu automatisables.

Par exemple, la transfection de vecteurs initialement hébergés
30 par des cellules de *E.coli* dans des cellules de *Streptomyces* nécessite classiquement une étape de récupération du vecteur recombinant contenu dans les cellules de *Escherichia coli*, et sa purification préalable à l'étape de transformation de protoplastes de *Streptomyces*. Il est
35 communément admis qu'une transfection d'un ensemble de 1000 clones

de *Escherichia coli* dans *Streptomyces* requiert l'obtention d'environ 8000 clones pour que chaque clone de *E. coli* ait une chance d'être représenté.

5 A l'inverse, une étape de transfection par conjugaison d'un vecteur hébergé par *E.coli* vers des cellules de *Streptomyces* nécessite le même nombre de clones de chacun des micro-organismes, l'étape de conjugaison ayant lieu " clone à clone " et ne comprenant en outre pas les difficultés techniques liées à l'étape de transfert de matériel génétique par transformation de protoplastes, par exemple en présence
10 de polyéthylène glycol.

Afin d'optimiser la construction de banque d'ADN chez *Streptomyces*, il a été mis au point selon l'invention, de nouveaux vecteurs conjuguatifs de type cosmide et de type BAC de nature à permettre une efficacité maximale de l'étape de conjugaison.

15 Notamment, les nouveaux vecteurs conjuguatifs selon l'invention ont été construits en plaçant un gène marqueur de sélection à l'extrémité de l'ADN du vecteur qui est transféré à la bactérie réceptrice en dernier lieu. Ce perfectionnement aux vecteurs conjuguatifs de l'état de la technique permet de ne sélectionner positivement que les bactéries
20 réceptrices ayant reçu la totalité de l'ADN du vecteur et, en conséquence, la totalité de l'ADN de l'insert d'intérêt.

Des cosmides conjuguatifs et intégratifs chez *Streptomyces* préférés selon l'invention sont les cosmides pOSV303, pOSV306 et pOSV307 décrits à l'exemple 5.

25 Selon un autre aspect, un procédé de préparation d'une collection de vecteurs recombinants selon l'invention est mis en oeuvre à l'aide d'un cosmide réplcatif à la fois chez *E.coli* et chez *Streptomyces*. Un tel cosmide est avantageusement le cosmide pOS 700R décrit à l'exemple 6.

30 Selon encore un autre aspect, le procédé ci-dessus peut être mis en oeuvre avec un cosmide réplcatif chez *E. coli* et *Streptomyces* et conjuguatif chez *Streptomyces*.

Un tel cosmide réplcatif et conjuguatif peut être obtenu à partir d'un cosmide réplcatif conforme à l'invention, par l'insertion d'une

origine de transfert appropriée, telle que RK2, comme décrit à l'exemple 5 pour la construction du vecteur pOSV303.

Selon un autre mode de réalisation avantageux du procédé de préparation d'une collection de vecteurs recombinants selon l'invention, on a recours à un vecteur de clonage et/ou d'expression de type BAC.

Selon un premier aspect, le vecteur du type BAC est intégratif et conjugatif chez *Streptomyces*.

De manière tout à fait préférée, un tel vecteur BAC intégratif et conjugatif chez *Streptomyces* est le vecteur BAC pOSV 403 décrit à l'exemple 8, ou encore les vecteurs BAC pMBD-1, pMBD-2, pMBD-3, pMBD-4, pMBD-5 et pMBD-6 décrits à l'exemple 15.

L'invention a en outre pour objet un vecteur recombinant caractérisé en ce qu'il est choisi parmi les vecteurs recombinants suivants:

a) un vecteur comprenant un acide nucléique constitutif d'une collection d'acides nucléiques selon l'invention;

b) un vecteur tel qu'obtenu selon un procédé éliminant tout recours à l'action d'une endonucléase de restriction sur le fragment d'ADN à insérer, tel que décrit précédemment.

De manière tout à fait préférée, l'invention est également relative à un vecteur choisi parmi les vecteurs suivants:

- le cosmide pOS700I;
- le cosmide pOSV303;
- le cosmide pOSV306;
- le cosmide pOSV307;
- le cosmide pOS700R;
- le vecteur BAC pOSV403;
- le vecteur BAC pMBD-1;
- le vecteur BAC pMBD-2;
- le vecteur BAC pMBD-3;
- le vecteur BAC pMBD-4;
- le vecteur BAC pMBD-5;
- le vecteur BAC pMBD-6.

L'invention est en outre relative à une collection de vecteurs recombinants tels qu'obtenus selon l'un quelconque des procédés selon l'invention.

5 **Procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression selon l'invention.**

Les techniques conventionnelles d'insertion d'ADN au sein d'un vecteur afin de préparer un vecteur de clonage et/ou d'expression
10 recombinant font classiquement appel à une première étape au cours de laquelle une endonucléase de restriction est incubée à la fois avec l'ADN à insérer et avec le vecteur récepteur créant ainsi des extrémités compatibles entre l'ADN à insérer et l'ADN du vecteur permettant l'assemblage des deux ADN avant une étape de ligation finale
15 permettant l'obtention du vecteur recombinant.

Toutefois, une telle technique conventionnelle présente des inconvénients notables, tout particulièrement lorsque est recherchée l'insertion d'acides nucléiques de grande taille dans un vecteur de clonage et/ou d'expression.

20 En effet, l'action préalable d'une enzyme de restriction sur les fragments d'ADN destinés à être insérés dans un vecteur est susceptible de réduire notablement la taille de cet ADN préalablement à son insertion dans le vecteur. Il va sans dire qu'une réduction significative de la taille de l'ADN préalablement à son insertion sur un vecteur est une
25 situation particulièrement défavorable lorsqu'est recherché le clonage de fragments d'ADN de grande taille susceptible de contenir l'ensemble des séquences codantes et, le cas échéant, également des séquences régulatrices, d'un opéron dont l'expression constitue une voie de biosynthèse complète d'un métabolite d'intérêt industriel, et tout
30 particulièrement d'un composé d'intérêt thérapeutique.

Pour remédier aux inconvénients des techniques de l'art antérieur, il a été mis au point selon l'invention deux procédés de préparation d'un vecteur recombinant de clonage et/ou d'expression qui ne nécessitent pas le recours à une endonucléase de restriction sur
35 l'ADN à insérer préalablement à son introduction au sein du vecteur. De

tels procédés sont en conséquence tout à fait adaptés au clonage de longs fragments d'ADN susceptibles de contenir, au moins partiellement, l'ensemble des séquences codantes et, le cas échéant, également des séquences régulatrices, d'un opéron complet responsable d'une voie de biosynthèse.

Selon un premier aspect, un procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression selon l'invention est caractérisé en ce que l'insertion d'un acide nucléique dans le vecteur de clonage et/ou d'expression, comprend les étapes suivantes:

- ouvrir le vecteur de clonage et/ou d'expression à un site de clonage choisi, à l'aide d'une endonucléase de restriction appropriée;

- ajouter un premier acide nucléique homopolymérique à l'extrémité 3' libre du vecteur ouvert;

- ajouter un second acide nucléique homopolymérique, de séquence complémentaire au premier acide nucléique homopolymérique, à l'extrémité 3' libre de l'acide nucléique à insérer dans le vecteur;

- assembler l'acide nucléique du vecteur et l'acide nucléique par hybridation du premier et du second acide nucléique homopolymérique de séquences complémentaires l'une de l'autre;

- refermer le vecteur par ligation.

Un tel procédé est décrit aux exemples 10 et 13 ci-après.

De manière avantageuse, le procédé ci-dessus peut comporter les caractéristiques suivantes, isolément ou en combinaison:

- le premier acide nucléique homopolymérique est de séquence poly(A) ou poly(T);

- le second acide nucléique homopolymérique est de séquence poly(T) ou poly(A).

De manière tout à fait préférée, les acides nucléiques homopolymériques ont une longueur comprise entre 25 et 100 bases nucléotidiques, préférentiellement entre 25 et 70 bases nucléotidiques.

Le procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression décrit ci-dessus est particulièrement adapté à la construction de banques d'ADN dans des vecteurs de type BAC. Ainsi, selon un mode de réalisation avantageux du procédé de préparation d'un vecteur recombinant décrit ci-dessus, ledit procédé est en outre caractérisé en ce que la taille de l'acide nucléique à insérer est d'au moins 100 kb, et préférentiellement d'au moins 200, 300, 400, 500 ou 600 kb.

Un tel procédé de préparation est donc particulièrement adapté à l'insertion des acides nucléiques contenus dans une collection d'acides nucléiques obtenus selon le procédé de l'invention.

Afin de permettre l'insertion de fragments d'ADN de grande taille dans des vecteurs de clonage et/ou d'expression, il a été mis au point selon l'invention, un second procédé ayant permis d'éliminer tout recours à l'action d'une endonucléase de restriction sur l'ADN destiné à être inséré au sein du vecteur.

Un tel procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression selon l'invention est caractérisé en ce que l'étape d'insertion d'un acide nucléique dans ledit vecteur de clonage et/ou d'expression comprend les étapes suivantes:

- création de bouts francs sur les extrémités de l'acide nucléique de la collection par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes;

- ouverture du vecteur de clonage et/ou d'expression à un site de clonage choisi à l'aide d'une endonucléase de restriction appropriée;

- addition d'adaptateurs oligonucléotidiques complémentaires ;

- création de bouts francs aux extrémités de l'acide nucléique du vecteur par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes, puis déphosphorylation des extrémités 5' afin de prévenir une recircularisation du vecteur;

- insertion de l'acide nucléique de la collection dans le vecteur par ligation.

De manière préférée, l'élimination des séquences 3' sortantes est réalisée à l'aide d'une exonucléase, telle que l'enzyme de Klenow.

De manière préférée, le remplissage des séquences 5' sortantes est réalisé à l'aide d'une polymérase, et de manière tout à fait préférée de la T4 polymérase, en présence des quatre nucléotides triphosphates.

Un procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes tel que décrit ci-dessus est particulièrement adapté à la construction de banques d'ADN à partir de vecteurs de type cosmide.

Un tel procédé d'obtention de vecteurs recombinants est décrit à l'exemple 12.

Dans un mode particulier de préparation d'un vecteur recombinant selon l'invention, des oligonucléotides comprenant un ou plusieurs sites de restriction rares sont ajoutés sur le vecteur au niveau du site de clonage de l'ADN à insérer, conformément à l'enseignement de l'exemple 10. Cet ajout d'oligonucléotides facilite la récupération ultérieure des inserts sans clivage de ces derniers.

CELLULES HOTES

Bien que tout type de cellules hôtes puisse être utilisé pour la transfection ou la transformation avec un acide nucléique ou un vecteur recombinant selon l'invention, notamment une cellule hôte procaryote ou eucaryote, on utilisera de préférence des cellules hôtes dont les

caractères physiologiques, biochimiques et génétiques sont bien caractérisés, facilement cultivables à grande échelle et dont les conditions de culture pour la production de métabolites soient bien connues.

5 De manière préférentielle, la cellule hôte réceptrice d'un acide nucléique ou d'un vecteur recombinant selon l'invention est phylogénétiquement proche des organismes donneurs contenus initialement dans l'échantillon de l'environnement desquels les acides nucléiques sont originaires.

10 De manière tout à fait préférée, une cellule hôte selon l'invention doit posséder un usage des codons similaire, ou du moins proche, des organismes donneurs présents initialement dans l'échantillon de l'environnement, tout particulièrement de l'échantillon de sol.

15 La taille des fragments d'ADN susceptible de porter les séquences nucléotidiques d'intérêt recherchées peut être variable. Ainsi, des enzymes codées par des gènes de taille moyenne de 1 kb pourront être exprimées à partir d'inserts de petite taille alors que l'expression de métabolites secondaires nécessiteront le maintien dans l'organisme hôte de fragments de taille bien supérieure, par exemple de 40 kb à plus de
20 100 kb, 200 kb, 300 kb, 400 kb ou 600 kb.

Ainsi, les cellules hôtes de *Escherichia coli* constituent un choix privilégié pour le clonage de grands fragments d'ADN.

25 De manière tout à fait préférée, on aura recours à l'utilisation de la souche de *Escherichia coli* désignée DH10B et décrite par Shizuya et al; (1992) pour laquelle des protocoles de clonage dans des vecteurs BAC ont été optimisés.

Toutefois, d'autres souches de *Escherichia coli* peuvent être avantageusement utilisées pour la construction d'une banque d'ADN selon l'invention, telles que les souches *E.coli* Sure, *E.coli* DH5 α , ou
30 encore *E.coli* 294 (ATCC N°31446).

En outre, la construction d'une banque d'ADN par transfection de cellules de *E.coli* avec des vecteurs recombinants selon l'invention est également possible, l'expression de gènes de divers procaryotes tels

que *Bacillus*, *Thermotoga*, *Corynebacterium*, *Lactobacillus* ou *Clostridium* ayant été décrite dans la demande PCT N°WO 99/20799.

De manière générale, des cellules hôtes de *E.coli* peuvent dans
5 tous les cas constituer des hôtes transitoires dans lesquels des vecteurs recombinants selon l'invention pourront être maintenus avec une grande efficacité, le matériel génétique pouvant être facilement manipulé et archivé et façon stable.

Dans le but d'exprimer la plus grande diversité moléculaire
10 possible, d'autres hôtes cellulaires pourront être également avantageusement mis en oeuvre tels que des cellules de *Bacillus*, *Pseudomonas*, *Streptomyces*, *Myxococcus*, *Aspergillus nidulans* ou encore *Neurospora crassa*.

Il a en outre été montré selon la présente invention, que des
15 cellules de *Streptomyces lividans* peuvent être utilisées avec succès et constituent des systèmes d'expression complémentaires à *Escherichia coli*.

Streptomyces lividans constitue un modèle pour l'étude de la génétique des *Streptomyces* et a également été utilisé comme hôte d'expression
20 hétérologue de nombreux métabolites secondaires. *Streptomyces lividans*, possède en commun avec d'autres actinomycètes tels que *Streptomyces coelicolor*, *Streptomyces griseus*, *Streptomyces fradiae*, ainsi que *Streptomyces griseochromogenes*, les molécules précurseurs et les systèmes de régulation nécessaires à l'expression de tout ou
25 partie des voies de biosynthèses complexes, telles que par exemple la voie de biosynthèse des polykétides ou encore la voie de biosynthèse des polypeptides non ribosomiques représentant des classes de molécules de structures très diverses.

Streptomyces lividans présente également l'avantage
30 d'accepter l'ADN étranger avec des efficacités de transformation élevées.

Ainsi, l'invention concerne aussi une cellule hôte recombinante comprenant un acide nucléique selon l'invention, constitutif d'une collection d'acides nucléiques préparée selon un procédé conforme à

l'invention, ou encore une cellule hôte recombinante comprenant un vecteur recombinant tel que défini précédemment.

Selon un premier aspect, il peut s'agir d'une cellule hôte recombinante d'origine procaryote ou eucaryote.

5 Avantageusement, une cellule recombinante selon l'invention est une bactérie, et de manière tout à fait préférée une bactérie choisie parmi *E.coli* et *Streptomyces*.

 Selon un autre aspect, une cellule hôte recombinante selon l'invention est caractérisée en ce qu'il s'agit d'une levure ou encore d'un
10 champignon filamenteux.

 L'invention a également trait à une collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutive de la collection comprenant un acide nucléique provenant d'une collection d'acides nucléiques réalisée conformément à un procédé de préparation d'une
15 collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes tel que décrit ci-dessus.

 L'invention est également relative à une collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutives de la collection comprenant un vecteur recombinant selon l'invention.

20 En raison de la grande taille des inserts il est nécessaire d'avoir une efficacité maximale de transformation. Dans ce but, une souche réceptrice de *Streptomyces lividans* exprimant l'intégrase de pSAM2 de façon constitutive afin de favoriser l'intégration site-spécifique du vecteur est préférée. Pour cela, le gène *int* sous contrôle d'un promoteur fort est
25 intégré dans le chromosome. La surproduction d'intégrase n'induit pas de phénomènes d'excision (Raynal et al., 1998).

 La production d'un nouveau métabolite à partir de l'insert pourrait être toxique pour *Streptomyces* si l'insert ne contient pas de gènes de résistance à l'antibiotique produit ou si ce gène est peu ou pas
30 exprimé. La capacité des différents gènes permettant à *Streptomyces ambofaciens* de résister à l'antibiotique qu'il produit est étudiée (Gourmelen et al., 1998; Pernodet et al., 1999). Certains de ces gènes codent des transporteurs de type ABC susceptibles de conférer un large spectre de résistance. Ces gènes peuvent être introduits et surexprimés
35 dans la souche hôte de *Streptomyces lividans*.

A l'inverse, une souche hypersensible aux antibiotiques peut être utilisée (Pernodet et al., 1996), afin de détecter dans la banque la présence de gènes de résistance. En effet, chez les micro-organismes producteurs d'antibiotique, ces gènes de résistance sont souvent associés aux gènes de la voie de biosynthèse de l'antibiotique. La sélection de clones résistants peut permettre d'effectuer simplement un premier tri avant les tests plus complexes de détection d'un nouveau métabolite produit par le clone.

10 **ISOLEMENT ET CARACTERISATION DE NOUVELLES SEQUENCES NUCLEOTIDIQUES CODANT POUR DES POLYKETIDES SYNTHASES.**

Selon l'invention, une collection de cellules hôtes recombinantes a été obtenue après transfection des cellules hôtes par une collection de vecteurs recombinants contenant chacun un insert d'acide nucléique provenant d'une collection d'acides nucléiques préparée conformément au procédé selon l'invention.

Plus précisément, les fragments d'ADN obtenus selon le procédé de l'invention dans lequel il est mis en oeuvre une étape d'extraction indirecte d'ADN des organismes contenus dans l'échantillon de sol ont été tout d'abord clonés dans le cosmide intégratif pOS700I.

L'étape d'insertion des fragments d'ADN dans le cosmide intégratif pOS700I a été réalisée selon le procédé de l'invention dans lequel des queues de polynucléotides homopolymériques poly(A) et poly(T) ont été ajoutées à l'extrémité 3' respectivement de l'acide nucléique du vecteur et des fragments d'ADN à insérer.

Les vecteurs recombinants ainsi construits ont été encapsidés dans des têtes de phage lambda et les phages obtenus ont été utilisés pour infecter des cellules de *E. coli* selon des techniques bien connues de l'homme du métier.

Une banque d'environ 5000 clones de *Escherichia coli* a été obtenue.

Cette banque de clones a été criblée avec des couples d'amorces spécifiques d'une séquence nucléotidique codant pour une

enzyme impliquée dans la voie de biosynthèse des polykétides, l'enzyme PKS de type I, aussi désignée β -kétosynthase.

On rappelle ici que les polykétides constituent une classe chimique d'une grande diversité structurale comprenant un nombre important de molécules d'intérêt pharmaceutique tels que la tylosine, la monensine, la ivermectine, l'érythromycine, la doxorubicine ou encore le FK506.

Les polykétides sont synthétisés par condensation de molécules d'acétate sous l'action d'enzymes appelées polykétide synthases (PKSs). Il existe deux types de polykétide synthases. Les polykétide synthases de type II sont impliquées en général dans la synthèse des antibiotiques aromatiques polycycliques et catalysent la condensation d'unités acétate de façon itérative.

Les polykétide synthases de type I sont impliquées dans la synthèse des polykétides macrocycliques ou macrolides et constituent des enzymes modulaires multifonctionnelles.

Compte-tenu de leur intérêt thérapeutique, il existe un besoin dans l'état de la technique d'isoler et de caractériser de nouvelles polykétides synthases qui peuvent être utilisées pour la production de nouveaux composés pharmaceutiques, notamment de nouveaux composés pharmaceutiques à activité antibiotique.

Le criblage de la banque de clones recombinants décrite ci-dessus à l'aide d'amorces PCR amplifiant sélectivement des séquences nucléotidiques codant pour des polykétide synthases de type I a permis d'identifier des clones recombinants contenant des inserts d'ADN comprenant une séquence nucléotidique codant pour de nouvelles polykétide synthases. Les séquences nucléotidiques codant pour ces nouvelles polykétides synthases sont référencées comme les séquences SEQ ID N°33 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120.

Un autre objet de l'invention consiste en un acide nucléique codant pour une nouvelle polykétide synthase I, caractérisé en ce qu'il comprend l'une des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120.

De préférence, un tel acide nucléique se présente sous une forme isolée et/ou purifiée.

L'invention concerne aussi un vecteur recombinant comprenant un polynucléotide comprenant l'une des séquences SEQ ID N°34 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120

L'invention a également trait à une cellule hôte recombinante
5 comprenant un acide nucléique choisi parmi les polynucléotides comprenant l'une des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 et SEQ ID N° 115 à SDEQ ID N°120 ainsi qu'à une cellule hôte recombinante comprenant un vecteur recombinant dans lequel est inséré un polynucléotide comprenant l'une des séquences
10 nucléotidiques SEQ ID N°34 à SEQ ID N°44 et SEQ ID N°115 à SEQ ID N°120.

Avantageusement, les vecteurs recombinants contenant un insert d'ADN codant pour une nouvelle polykétide synthase de type I selon l'invention sont des vecteurs de clonage et d'expression.

15 De préférence, une cellule hôte recombinante telle que décrite ci-dessus est une bactérie, une levure ou encore un champignon filamenteux.

Les séquences en acides aminés de nouvelles polykétide synthases provenant d'organismes contenus dans un échantillon de sol
20 ont été déduites des séquences nucléotidiques SEQ ID N°34 à SEQ ID N°44 ET SEQ ID N° 115 à SEQ ID N°120 ci-dessus. Il s'agit des polypeptides comprenant l'une des séquences en acides aminés SEQ ID N°48 à SEQ ID N°59 et SEQ ID N° 121 à 126.

L'invention concerne encore de nouvelles polykétides synthases comprenant une séquence en acides aminés choisie parmi
25 les séquences SEQ ID N°48 à SEQ ID N°59 et SEQ ID N° 121 à SEQ ID N°126.

Fait également partie de l'invention la séquence nucléotidique SEQ ID N°114 qui comprend six cadres ouverts de lecture qui codent
30 respectivement les polypeptides de séquences SEQ ID N°121 à SEQ ID N°126.

Fait également partie de l'invention la séquence nucléotidique SEQ ID N°113 du cosmide a26G1, qui contient la séquence complémentaire de la séquence SEQ ID N°114.

On a aussi extrait et amplifié selon l'invention de l'ADN génomique provenant de souches bactériennes pures, telles que *Streptomyces coelicolor* (ATCC N°101.478), *Streptomyces ambofaciens* (NRRL N°2.420), *Streptomyces lactamandurans* (ATCC N°27.382),
5 *Streptomyces rimosus* (ATCC N°109.610), *Bacillus subtilis* (ATCC N°6633) ou encore *Bacillus licheniformis* et *Saccharopolyspora erythrea*.

Une amplification par PCR de l'ADN de chacune des souches bactériennes décrites ci-dessus a été effectuée à l'aide des couples d'amorces spécifiques des séquences nucléiques de polykétide
10 synthase de type I.

De nouveaux gènes de polykétide synthases de type I bactériennes ont ainsi pu être isolés et caractérisés. Il s'agit des séquences nucléiques de séquences SEQ ID N°30 à SEQ ID N°32.

L'invention a donc en outre pour objet des séquences
15 nucléotidiques codant pour de nouvelles polykétides synthases de type I choisies parmi les polynucléotides comprenant l'une des séquences nucléotidiques SEQ ID N°30 à SEQ ID N°32.

Font également partie de l'invention des vecteurs recombinants comprenant les séquences nucléotidiques codant pour de nouvelles
20 polykétides synthases de type I définies ci-dessus.

L'invention concerne aussi des cellules hôtes recombinantes caractérisées en ce qu'elles contiennent un acide nucléique codant pour une nouvelle polykétide synthase de type I comprenant une séquence nucléotidique choisie parmi les séquences SEQ ID N°30 à SEQ ID N°32
25 ainsi que des cellules hôtes recombinantes comprenant un vecteur recombinant tel que défini ci-dessus.

L'invention a également pour objet des polypeptides codés par des séquences comprenant les acides nucléiques SEQ ID N° 30 à 32, et plus précisément des polypeptides comprenant les séquences d'acides
30 aminés SEQ ID N° 47 à SEQ ID N° 50.

L'invention a en outre pour objet un procédé de production d'une polykétide synthase de type I selon l'invention, ledit procédé de production comprenant les étapes suivantes:

- obtention d'une cellule hôte recombinante comprenant un acide nucléique codant pour une polykétide synthase de type I comprenant une séquence nucléotidique choisie parmi les séquences SEQ ID N°33 à SEQ ID N°44, SEQ ID N°30 à SEQ ID N°32 et SEQ ID N°115 à SEQ ID N°120;

- culture des cellules hôtes recombinantes dans un milieu de culture approprié;

- récupération et, le cas échéant, purification de la polykétide synthase de type I à partir du surnageant de culture ou du lysat cellulaire.

Les nouvelles polykétide synthases de type I obtenues selon le procédé décrit ci-dessus peuvent être caractérisées par fixation sur une colonne de chromatographie d'immuno-affinité sur laquelle des anticorps reconnaissant ces polykétides synthases ont été préalablement immobilisés.

Les polykétide synthases de type I selon l'invention, et plus particulièrement les polykétide synthases recombinantes décrites ci-dessus peuvent être aussi purifiées par des techniques de chromatographie liquide à haute performance (HPLC), telles que par exemple des techniques de chromatographie en phase inverse ou de chromatographie d'échanges d'anions ou de cations, bien connues de l'homme du métier.

Les polykétide synthases, recombinantes ou non recombinantes, selon l'invention peuvent être utilisées pour la préparation d'anticorps.

Selon un autre aspect, l'invention a donc encore pour objet un anticorps reconnaissant spécifiquement une polykétide synthase de type I selon l'invention ou un fragment peptidique d'une telle polykétide synthase.

Les anticorps selon l'invention peuvent être monoclonaux ou polyclonaux. Les anticorps monoclonaux peuvent être préparés à partir

de cellules d'hybridome selon la technique décrite par KOHLER et MILSTEIN C. (1975), Nature, Vol.256:495.

Les anticorps polyclonaux peuvent être préparés par immunisation d'un mammifère, en particulier des souris, des rats ou des lapins avec une polykétide synthase de type I selon l'invention, le cas échéant en présence d'un composé adjuvant de l'immunité, tels que l'adjuvant complet de Freund, l'adjuvant incomplet de Freund, l'hydroxyde d'aluminium ou encore un composé de la famille des muramyl peptides.

Constituent également des " anticorps " au sens de la présente invention, les fragments d'anticorps tels que les fragments Fab, Fab', F(ab')₂, ou encore les fragments d'anticorps simple chaîne contenant la partie variable (ScFv) décrits par MARTINEAU et al. (1998) J. Mol. Biol., Vol.280 (1):117-127 ou encore dans le brevet US 4,946,778, ainsi que les anticorps humanisés décrits par REINMANN KA et al. (1997), AIDS Res. Hum. Retroviruses, vol.13(11):933-943 ou par LEGER O.J et al. (1997), Hum. Antibodies, vol.8 (1): 3-16.

Les préparations d'anticorps selon l'invention sont utiles notamment dans des tests immunologiques qualitatifs ou quantitatifs visant, soit à simplement détecter la présence d'une polykétide synthase de type I selon l'invention, soit à quantifier la quantité de cette polykétide synthase, par exemple dans le surnageant de culture ou le lysat cellulaire d'une souche bactérienne susceptible de produire une telle enzyme.

Un autre objet de l'invention consiste en un procédé de détection d'une polykétide synthase de type I selon l'invention ou un fragment peptidique de cette enzyme, dans un échantillon, ledit procédé comprenant les étapes de :

a) mettre en contact un anticorps selon l'invention avec l'échantillon à tester;

b) détecter le complexe antigène/anticorps éventuellement formé.

L'invention est également relative à un kit ou nécessaire de détection d'une polykétide synthase de type I selon l'invention dans un échantillon, comprenant :

- a) un anticorps selon l'invention;
- 5 b) le cas échéant, des réactifs nécessaires à la détection du complexe antigène/anticorps éventuellement formé.

Un anticorps dirigé contre une polykétide synthase de type I selon l'invention peut être marqué à l'aide d'un marqueur détectable isotopique ou non isotopique, selon des procédés bien connus de
10 l'homme du métier.

Le criblage d'une banque d'ADN selon l'invention à l'aide d'une paire d'amorces hybridant avec des séquences cibles dont la présence est recherchée, telles que des séquences de la voie de biosynthèse de la puromycine, des séquences du gène *linA* impliquées dans la
15 biodégradation du lindane ou encore des séquences codant pour des polykétides synthases de type I ont été détaillées ci-avant.

L'invention a donc pour objet un procédé de détection d'un acide nucléique de séquence nucléotidique déterminée, ou de séquence nucléotidique structuralement apparentée à une séquence nucléotidique
20 déterminée, dans une collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- mettre en contact la collection de cellules hôtes recombinantes avec un couple d'amorces hybridant avec la séquence nucléotidique déterminée ou hybridant avec la séquence nucléotidique
25 structurellement apparentée à une séquence nucléotidique déterminée;
- réaliser au moins trois cycles d'amplification ;
- détecter l'acide nucléique éventuellement amplifié.

Pour les conditions d'amplification appropriées en fonction des
30 séquences cibles recherchées, l'homme du métier pourra se référer avantageusement aux exemples ci-dessous.

Selon un autre aspect, l'invention concerne aussi un procédé de détection d'un acide nucléique, de séquences nucléotidiques déterminées, ou de séquences nucléotidiques structurellement
35 apparentées à une séquence nucléotidique déterminée, dans une

collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- mettre en contact la collection de cellules hôtes recombinantes avec une sonde hybridant avec la séquence nucléotidique déterminée ou hybridant avec une séquence nucléotidique structurellement apparentée à la séquence nucléotidique déterminée;
- détecter l'hybride éventuellement formé entre la sonde et les acides nucléiques compris dans les vecteurs de la collection.

10 Pour effectuer le criblage d'une banque d'ADN selon l'invention en vue de détecter la présence d'une séquence nucléotidique codant pour un polypeptide capable de dégrader le lindane, on a détecté les clones recombinants d'intérêt par leur phénotype correspondant à leur capacité à dégrader le lindane. Dans ce but, les clones isolés et/ou des
15 ensembles de clones de la banque d'ADN préparée ont été mis en culture dans un milieu de culture en présence de lindane et la dégradation du lindane a été observée par la formation d'un halo trouble dans l'environnement immédiat des cellules.

L'invention concerne aussi un procédé pour identifier la
20 production d'un composé d'intérêt par une ou plusieurs cellules hôtes recombinantes dans une collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié;
- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules recombinantes cultivées.

L'invention a en outre pour objet un procédé pour sélectionner une cellule hôte recombinante produisant un composé d'intérêt dans une
30 collection de cellules hôtes recombinantes selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes:

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié;

- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules hôtes recombinantes cultivées;

- sélection des cellules hôtes recombinantes produisant le composé d'intérêt.

L'invention concerne encore un procédé pour la production d'un composé d'intérêt caractérisé en ce qu'il comprend les étapes suivantes:

- cultiver une cellule hôte recombinante sélectionnée selon le procédé décrit ci-dessus;

- récupérer et, le cas échéant, purifier, le composé produit par ladite cellule hôte recombinante.

L'invention est également relative à un composé d'intérêt caractérisé en ce qu'il est obtenu selon le procédé ci-dessus décrit.

Un composé d'intérêt selon l'invention peut consister en un polykétide produit grâce à l'expression d'au moins une séquence nucléotidique comprenant une séquence choisie parmi les séquences SEQ ID N°33 à 44, SEQ ID N°30 à 32 et SEQ ID N°115 à SEQ ID N°120.

L'invention concerne encore une composition comprenant un polykétide produit grâce à l'expression d'au moins une séquence nucléotidique comprenant une séquence choisie parmi les séquences SEQ ID N°33 à SEQ ID N°44, SEQ ID N°30 à SEQ ID N°32, et SEQ ID N°115 à SEQ ID N°120.

Un polykétide produit grâce à l'expression d'au moins une séquence nucléotidique ci-dessus est préférentiellement le produit de l'activité de plusieurs séquences codantes incluses au sein d'un opéron fonctionnel dont les produits de traduction sont les différentes enzymes nécessaires à la synthèse d'un polykétide, l'une des séquences ci-dessus étant comprise et exprimée dans ledit opéron. Un tel opéron comprenant une séquence d'acide nucléique selon l'invention codant pour une polykétide synthase peut être construit par exemple selon l'enseignement de Borchert et al. (1992).

L'invention est encore relative à une composition pharmaceutique comprenant une quantité pharmacologiquement active

d'un polykétide selon l'invention, le cas échéant en association avec un véhicule pharmaceutiquement compatible.

De telles compositions pharmaceutiques seront avantageusement adaptées pour l'administration, par exemple par voie parentérale, d'une quantité d'un polykétide synthétisé par une polykétide synthase de type I selon l'invention allant de $1\mu\text{g/kg}$ par jour à 10 mg/kg par jour, de préférence au moins $0,01\text{ mg/kg}$ par jour et de manière tout à fait préférée entre $0,01$ et 1 mg/kg par jour.

Les compositions pharmaceutiques selon l'invention peuvent être indifféremment administrées par voie orale, rectale, parentérale, intraveineuse, sous-cutanée ou encore intradermique.

L'invention concerne aussi l'utilisation d'un polykétide obtenu grâce à l'expression d'une polykétide synthase de type I selon l'invention pour la fabrication d'un médicament, en particulier d'un médicament à activité antibiotique.

L'invention sera en outre illustrée, sans pour autant être limitée, par les figures et les exemples ci-après.

La figure 1 illustre le schéma des différentes étapes de lyse effectuées selon les protocoles 1, 2, 3n 4a, 4b, 5a, et 5b décrits à l'exemple 1.

La Figure 2 illustre une électrophorèse sur gel d'agarose 0.8% des ADN extraits à partir de 300 mg du sol n°3 (Côte St André) après différents traitements de lyse (protocoles 1 à 5, cf. Fig. 1). M : marqueur de poids moléculaire de phage lambda

La Figure 3 illustre la proportion de différents genres d'actinomycètes cultivés à la suite des traitements 1 à 5 (cf. Fig. 1). Le nombre d'ufc (unité formant colonie) a été déterminé sur un milieu sélectif pour ce groupe de bactéries. Un nombre total d'environ 400 colonies a été analysé.

La Figure 4 illustre la récupération d'ADN de phage lambda digéré par *HindIII* additionné dans les sols à différentes concentrations avant (G) ou après (G*) broyage. Les traitements T (chocs thermiques)

et S (sonication) sont des traitements additionnels de lyse. La quantification a été réalisée par analyse au phospho-imageur après hybridation en dot-blot. Un échantillon de chaque sol a été utilisé pour chaque concentration de phage lambda ajouté. Les caractéristiques des sol sont reproduites dans le tableau 1. Les échantillons correspondant à 10 et 15 µg d'ADN ajouté n'ont pas été traités.

La Figure 5 illustre l'amplification par PCR des ADN extraits à partir de sol n°3 selon les protocoles 1, 2, 3, 5a et 5b. Les amorces FGPS 122 et FGPS 350 (tableau 2) ont été utilisés afin de cibler *Streptosporangium spp.* indigènes. Les extraits d'ADN ont été utilisés non dilués ou dilués au 1/10^{ème} et 1/100^{ème}. M : marqueur de poids moléculaires 123 pb (Gibco BRL), C : contrôle d'amplification sans ADN.

La Figure 6 illustre les quantités d'ADN extrait après inoculation de spores (a) ou de mycélium (b) de *S. lividans* OS48.3 inoculés dans les sols à différentes concentrations. La quantités de mycélium ajoutée dans le sol correspond au nombre de spores inoculées dans le milieu de germination. Environ 50% des spores ont germé, le nombre de cellules ou de génomes contenues dans les hyphes des spores germées n'a pas été déterminé. Les quantités de spores et de mycélium inoculées ne sont donc pas directement comparables. Le protocole d'extraction a été mené selon le protocole 6 (cf. section matériel et méthodes). Le symbole (') indique que de l'ARN a été inclus dans le tampon d'extraction. L'ADN cible a été amplifié par PCR avec les amorces FGPS 516 et FGPS 517, la quantification a été réalisée par phosphoimageur après hybridation en dot blot en utilisant le sonde FGPS 518. Un échantillon de chaque sol a été utilisé pour chaque concentration d'hyphes ou de spores. Les caractéristiques des sols sont décrites dans le tableau 1.

La figure 7 représente l'arbre phylogénétique obtenu par l'algorithme de Neighbour Joining , positionnant les séquences d'ADNr 16S contenues dans la banque d'ADN du sol, par rapport à des bactéries de références cultivées.

En grisé: les séquences issues des pools de clones de la banque.

Les valeurs de bootstrap sont indiquées au niveau des noeuds, après rééchantillonnage de 100 répétitions. La barre d'échelle indique le nombre de substitutions par site. Le numéro d'accès des séquences dans la base de données Genbank est indiqué entre parenthèses.

La figure 8 représente un schéma du vecteur pOSint 1.

La figure 9 représente un schéma du vecteur pWED1.

La figure 10 représente un schéma du vecteur pWE15 (ATCC N° 37503).

La figure 11 représente un schéma du vecteur pOS 700I.

La figure 12 représente un schéma du vecteur pOSV010.

La figure 13 représente le fragment contenant un site "cos" inséré dans le plasmide pOSV010 au cours de la construction du vecteur pOSV 303.

La figure 14 représente un schéma du vecteur pOSV 303.

La figure 15 représente un schéma du vecteur pE116.

La figure 16 représente un schéma du vecteur pOS 700 R.

La figure 17 représente un schéma du vecteur pOSV 001.

La figure 18 représente le schéma du vecteur pOSV 002.

La figure 19 représente un schéma du vecteur pOSV 014.

La figure 20 représente un schéma du vecteur pBAC 11.

La figure 21 représente un schéma du vecteur pOSV 403.

La figure 22 représente les gels d'électrophorèse d'ADN de la banque après digestion par les enzymes BamHI et DraI des clones positifs de la banque criblée avec les oligonucléotides PKS-I.

La figure 23 illustre la production de puromycine par les recombinants de *S. lividans* comparée à la production de la souche sauvage *S. alboniger*.

10

La Figure 24 illustre l'alignement de PKSs du sol avec les sites actifs conservés d'autres PKSs. Les références pour chaque peptide sont indiquées. Les domaines bêta-kétoacyl synthase ont été alignés en utilisant le programme PILEUP de GCG (Wisconsin Package Version 9.1, Genetics Computer Group, Madison, Wisc).

15

La Figure 25 illustre la construction d'un cosmide intégratif conjugatif.

20

La Figure 26 illustre la construction d'un BAC intégratif conjugatif.

La figure 27 illustre le schéma de construction du vecteur pOSV 308.

25

La figure 28 illustre le schéma de construction du vecteur pOSV306.

La figure 29 illustre le schéma de construction du vecteur pOSV307.

30

La figure 30 illustre le schéma de construction du vecteur PMBD-1.

La figure 31 présente une carte détaillée du plasmide pMBD-2 ainsi qu'un schéma de construction du vecteur pMBD-3.

La figure 32 illustre une carte détaillée du plasmide pMBD-4.

La figure 33 illustre le schéma de construction du plasmide pMBD-5 à partir du plasmide pMBD-1.

La figure 34 illustre la carte détaillée du vecteur pBTP-3.

La figure 35 illustre le schéma de construction du vecteur pMBD-6 à partir du vecteur pMBD-1.

La figure 36 illustre la carte du cosmide a26G1 dont l'insert d'ADN contient des cadres ouverts de lecture codant pour plusieurs polykétides synthase.

La figure 37 est un schéma représentant l'insert d'ADN (brin +) du cosmide a26G1, sur lequel sont positionnés les différents cadres de lecture codant pour plusieurs polykétides synthase.

EXEMPLES:

EXEMPLE 1: Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes, comprenant une étape d'extraction directe d'ADN à partir de l'échantillon de sol.

1. MATERIEL ET METHODES

1.1 SOLS: Les caractéristiques des six sols utilisés dans cette étude sont listées dans le tableau 1.

La teneur en argile et en matière organique va respectivement de 9 à 47% et de 1,7 à 4,7%, le pH variant de 4,3 à 5,8.

Des échantillons de sol ont été collectés à partir de la couche superficielle de 5 à 10 cm de profondeur. Toutes les racines visibles ont

été éliminées et les sols ont été conservés à 4°C pendant quelques jours si nécessaire, après quoi ils ont été séchés pendant 24 heures à la température ambiante et tamisés (taille moyenne de maille 2 mm) avant d'être conservés jusqu'à plusieurs mois à 4°C.

5

1.2 SOUCHES BACTERIENNES ET CONDITIONS DE CULTURE:

L'ADN extracellulaire ainsi que les souches bactériennes fournissant des cellules végétatives, des spores ou des hyphae, utilisées pour innoculer les échantillons de sol, ont été choisies de telle sorte que leur présence
10 puisse être suivie spécifiquement.

Afin d'obtenir de grandes quantités d'ADN extracellulaire, la souche lysogénique de *E.coli* 1192 Hfr P4X (metB), contenant le phage lambda CI857 Sam7, a été cultivée sur milieu Luria-Bertani (LB) pendant deux heures à 30°C, puis 30 minutes à 40°C, puis 3 heures à 37°C.

15 L'ADN du phage lambda a été extrait selon la technique décrite par SAMBROOK J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor N.Y.

La souche avirulente de *Bacillus anthracis* (STERNE 7700) a été utilisée comme inoculum de cellules bactériennes. *Bacillus anthracis*
20 a été multiplié sur un bouillon de culture de type " trypticase soy broth " (TSB) (Biomérieux, Lyon, France) pendant environ 6 heures, en vérifiant que la DO₆₀₀ soit maintenue en dessous de 0,6. Ces conditions permettent le développement des cellules végétatives sans formation de spores (Patra et al., (1996), FEMS Immunol. Medical Microbiology, vol.15:223-231.). Les spores de *Streptomyces lividans* OS48.3 (CLERC-
25 BARDIN et al. non publié) ont été éliminées mécaniquement des cultures de l'organisme sur un milieu R2YE (HOPWOOD et al., (1985), Genetic Manipulation of Streptomyces-A Laboratory Manual. The John Innes Foundation , Norwich ,United Kingdom). Les hyphae de *S.lividans*
30 OS48.3 ont été obtenus à partir des spores en pré-germination, car l'on s'attendait à ce que l'utilisation de hyphae courtes minimise la rupture et la perte subséquente d'ADN. Les spores ont été mises en suspension dans du tampon TES (Acide N-Tris [hydroxyméthyl]méthyl-2-aminoéthanesulfonique ; Sigma-Aldrich Chimie, France) (0,05M; pH 8)
35 (Holben WE et al., (1988), APPL. Environ. Microbiol. vol.54:703-711,

puis ont été soumises à un choc thermique (50°C pendant 10 minutes suivi d'un refroidissement sous un courant d'eau froide puis ajoutées à un volume égal de milieu de pré-germination (extrait de levure 1%, casaminoacides 1% CaCl₂ 0,01 M).

5 La solution a été incubée à 37°C sur un agitateur. La proportion de spores germées a été estimée à environ 50%, en accord avec les résultats de HOPWOOD et al. (1985). Après centrifugation, les culots ont été resuspendus dans du tampon TES, ajoutés à 3% de milieu TSB, et incubés à 37°C jusqu'à l'obtention d'une DO₄₅₀ de 0,15 (HOPWOOD et
10 al. , (1985)). *Streptomyces hygroscopicus* SWN 736 et *Streptosporangium fragile* AC1296 (Institute Pushino, Moscou) ont été cultivés selon des techniques décrites par HICKEY et TRESNER (1952).

L'ADN des spores et des hyphae de *S. Lividans* a été extrait à partir des cultures pures selon le protocole de lyse 6 décrit ci-dessous
15 (excepté qu'aucun broyage n'a été réalisé), tandis que les spores de *S. hygroscopicus* et de *S. fragile* ont été extraites par lyse chimique/enzymatique (Hintermann et al., 1981).

1.3 CHOIX DU TAMPON D'EXTRACTION: Un tampon TENP (50 mM
20 Tris, 20 mM EDTA, 100 mM NaCl, 1% pds/vol de polyvinylpolypyrrolidone développé par PICARD (1992) a été utilisé. Des tampons similaires ont été ultérieurement utilisés par d'autres auteurs (CLEGG et al., 1997; KUSKE et al., 1998; ZHOU et al., 1996).

Le Tris et l'EDTA protègent l'ADN de l'activité nucléase, le NaCl
25 apporte un effet dispersant et la PVPP absorbe les acides humiques et les autres composés phénoliques (HOLBEN et al. (1988); PICARD et al., (1992).

Dans cette étude, l'efficacité d'extraction de ce tampon a été évaluée à différents pH (6,0 - 10,0) en utilisant 20 sols différents ayant
30 une gamme de pH de 5,8 à 8,3 et une teneur en matière organique entre 0,2 et 6,3%. Ces vingt sols (les autres caractéristiques ne sont pas indiquées) ont été utilisés uniquement dans cette expérience. La quantité d'ADN a été déterminée de manière colorimétrique comme décrit par RICHARD (1974), et détaillé ci-après.

1.4 PROTOCOLE DE LYE *IN SITU* ET D'EXTRACTION D'ADN:

Plusieurs protocoles utilisant un nombre croissant d'étapes ont été testés afin d'évaluer l'efficacité de différentes techniques pour lyser les microbes du sol *in situ*. Pour ces expériences, la microflore indigène du sol a été ciblée dans six sols. Des expériences additionnelles ont été conduites afin d'étudier les effets des traitements de lyse sur l'ADN libéré, en analysant les quantités et la qualité d'ADN récupéré provenant d'un ADN de phage lambda préalablement additionné aux sols.

Une fois qu'un protocole optimisé (désigné protocole 6) a été développé, ce protocole a été utilisé pour quantifier l'ADN provenant d'*Actinomycètes* indigènes et d'ADN provenant de bactéries Gram-positives inoculées dans les sols sélectionnés. Dans tous les cas, les échantillons de sol ont été séchés et passés au tamis comme décrit ci-dessus.

Après broyage, 0,5 ml de tampon TENP ont été ajoutés à 200 mg poids sec de sol excepté pour le protocole 1 dans lequel le tampon a été ajouté à un sol non broyé).

Pour les divers traitements de lyse (voir ci-dessous), les suspensions de sol ont été passées au Vortex pendant dix minutes et centrifugées (4000 g pendant cinq minutes), après quoi une fraction aliquote (25 µl) du surnageant a été analysée par électrophorèse sur gel (0,8% d'agarose).

Une autre fraction aliquote du surnageant représentant un volume connu, généralement 350 µl, a été précipitée avec de l'isopropanol.

Cinq fractions aliquotes (représentant de l'ADN dérivé de 1 g de sol) ont été réunies et resuspendues dans 100 µl d'un tampon TE stérile (10 mM Tris, 1 mM EDTA, pH 8,0) avant purification (protocole D, voir ci-dessous) et quantification, soit par hybridation (Dot Blot) de l'ADN total, soit par hybridation (Dot Blot) des produits d'amplification PCR (voir ci-dessous).

Les signaux d'hybridation ont été quantifiés par imagerie par phosphorescence (technique de "phospho-imaging" voir ci-dessous).

1.5 EVALUATION DES METHODES DE LYSE CELLULAIRE *IN SITU*:

La qualité et la quantité de l'ADN extrait après un nombre croissant d'étapes de traitement de lyse (protocole 2-5b) ont été comparées à celles de l'ADN extracellulaire obtenu après lavage du sol avec un
5 tampon d'extraction (protocole 1; voir aussi figure 1).

Protocole 1: Pas de traitement de lyse.

Le tampon TENP a été ajouté à un sol non broyé, une étape
10 d'extraction d'ADN a été réalisée comme décrit ci-dessus.

Protocole 2. Broyage du sol suivi d'une extraction d'ADN.

Deux types de dispositifs différents ont été utilisés pour le
15 broyage du sol.

Afin de comparer leur efficacité respective, 5g de sol sec ont été broyés pendant 30 secondes dans un broyeur contenant des anneaux de tungstène, ou pendant des temps variés jusqu'à 60 minutes dans un broyeur de sol contenant un mortier et des billes en agate (20
20 mm de diamètre).

Le tampon TENP est ensuite ajouté et l'ADN est extrait comme décrit ci-dessus.

Les résultats d'électrophorèse sur gel ont montré qu'un broyage de 40 minutes en utilisant des billes en agate étaient nécessaires afin
25 d'obtenir des quantités d'ADN extraits équivalentes à celles obtenues après 30 secondes de broyage en utilisant des anneaux de tungstène.

La distribution de taille des fragments d'ADN est similaire quelle que soit la méthode employée.

Ainsi, ces traitements ont été considérés comme équivalents et
30 celui qui sera utilisé dans les protocoles décrits ci-dessous ne sera en conséquence pas spécifié.

Dans les protocoles 3 à 5, l'efficacité de plusieurs autres traitements de lyse ultérieure au broyage du sol a été testée, soit séparément, soit dans différentes combinaisons.

Protocole 3:

Ce protocole est identique au protocole 2 , sauf qu'il comprend une étape d'homogénéisation à l'aide d'un mixeur de type Ultraturrax (Janker et Kunkel, IKA Labortechnik, Allemagne) réglé à la moitié de la vitesse maximale pendant 5 minutes.

PROTOCOLES 4a et 4b:

Ces protocoles sont identiques au protocole 3 à l'exception d'une étape additionnelle de sonication.

Deux types de dispositifs sonicateurs ont été comparés : un sonicateur à micropointe de titane (600W Vibracell Ultrasonicator, Bioblock, Illkirch, France) (Protocole 4a) et un sonicateur de type Cup Horn (protocole 4b).

La micropointe Vibracell produisant des ultrasons est en contact direct avec la solution de sol.

En ce qui concerne le dispositif de type Cup Horn, la solution de sol est conservée dans des tubes qui sont placés dans un bain d'eau à travers lequel passent les ultrasons.

Des expériences préliminaires ont été réalisées afin de déterminer les conditions optimales pour les deux sonicateurs (résultats non présentés).

Le meilleur compromis, en terme de quantité d'ADN extrait et de taille de fragments, consiste en une sonication avec la micropointe de titane et le sonicateur de type Cup Horn respectivement pendant 7 et 10 minutes, en réglant la puissance à 15 W et avec des cycles actifs à 50%.

Protocoles 5a et 5b:

Après sonication avec une micropointe de titane ou un dispositif de type Cup Horn (respectivement protocoles 4a et 4b) du lysozyme et de l'achromopeptidase ont été ajoutés, chacune des enzymes à une concentration finale de 0,3 mg/ml.

Les suspensions de sol ont été incubées pendant 30 minutes à 37°C, après quoi du lauryl sulfate à une concentration finale de 1 % a été ajouté, puis des suspensions ont été incubées pendant 1 heure à 60°C avant centrifugation et précipitation comme décrit ci-dessus.

5 En plus des protocoles décrits ci-dessus, l'effet de la sonication (Cup Horn, voir protocole 4b) et de chocs thermiques (30 secondes dans l'azote liquide suivi de trois minutes dans l'eau bouillante, les traitements étant répétés trois fois) sur l'ADN de phage lambda digéré par HindIII préalablement ajouté au sol ont été examinés (voir ci-après).

10 Des chocs thermiques ont été suggérés dans l'état de la technique comme des moyens de lyse cellulaire *in situ* (PICARD et al. (1992)). Cependant, du fait qu'un tel traitement a un effet préjudiciable sur l'ADN libre (voir la section résultats) il n'a pas été inclus dans les protocoles décrits ci-dessus.

15

PROTOCOLE OPTIMISE

Après évaluation des différents traitements de lyse, un protocole optimisé a été défini, désigné protocole 6. Le protocole 6 est
20 identique au protocole 5b excepté que, avant la sonication, les suspensions de sol sont soumises à un traitement par Vortex puis agitées par rotation sur une roue pendant deux heures avant d'être congelées à - 20°C.

Après décongélation, les suspensions de sol sont passées au
25 Vortex pendant 10 minutes avant sonication. Le protocole 6 a été utilisé dans les expériences dans lesquelles les sols ont été ensemencés avec des cellules bactériennes ainsi que dans les expériences dans lesquelles les actinomycètes indigènes ont été quantifiés (voir ci-dessous).

30 **1.6 COMPTAGE AU MICROSCOPE:** L'efficacité du broyage du sol comme méthode pour lyser des cellules bactériennes a été examinée au microscope.

5g de sol brut séché ont été mélangés dans un dispositif de type Waring Blender avec 50 ml d'eau stérilisée ultrapure pendant 1,5
35 minutes; simultanément, 1g (poids sec) de sol broyé (protocole n°2) a

été mis en suspension dans 10 ml par agitation pendant 10 minutes. Les suspensions de sol ont fait l'objet de dilutions en séries et de l'acridine orange a été ajoutée à une concentration finale de 0,001%.

Après 2 minutes, les suspensions ont été filtrées à travers une
5 membrane de marque NUCLEOPORE de type 0,2 µm black. Chaque filtre a été rincé avec de l'eau stérile lysée, traitée avec 1 ml d'isopropanol pendant 1 minute afin de fixer les cellules bactériennes, puis rincé de nouveau.

Les cellules bactériennes ont été comptées à l'aide d'un
10 microscope à épifluorescence du type Zeiss Universal avec un objectif 100x. Pour chacun des types de sol, trois filtres ont été comptés, et au moins 200 cellules ont été comptées sur chacun des filtres.

1.7 NUMERATION DES ACTINOMYCETES CULTIVABLES ET
15 **NOMBRE TOTAL D'UNITES FORMANT COLONIES (CFU):** Les actinomycètes ayant survécu aux traitements de lyse (protocoles 1-5) ont été examinés spécifiquement avec le sol n°3 (Côte Saint André, voir tableau 1).

Après une dilution de 10 fois d'une solution d'extrait de levure
20 (6% poids/volume) et de SDS (0,05%) afin d'induire la germination (Hayakawa et al. (1988)), les suspensions de sol ont été diluées en séries dans de l'eau stérile, incubées à 40°C pendant 20 minutes et ensemencées sur du milieu HV (HAYAKAWA et al., 1987).

Le milieu HV a été additionné de actidione (50 mg/l) et de
25 nystatine (50 mg/ml).

Les colonies d'actinomycètes ont été comptées après incubation pendant 15 jours à 28°C.

Au total, environ 400 colonies ont été examinées. L'identification a été réalisée sur la base des caractéristiques
30 morphologiques macro-et microscopiques ainsi que sur l'analyse de la teneur en acide diaminopimélique des isolats (SHIRLING et al., 1966); STANECK et al., 1974; WILLIAMS et al., 1993).

La quantité totale de bactéries cultivables (CFU totales) a été également déterminée pour chacun des protocoles de lyse 1 à 5. Les
35 suspensions de sol ont été diluées en série et ensemencées en triple sur

un milieu agar Bennett (WAKSMAN et al., 1961) additionné de nystatine et d'actidione (chacune à 50 mg/l).

Chaque boîte de Pétri a été couverte d'un filtre de nitrate de cellulose (Millipore) et incubée pendant trois jours à 28°C. Après la
5 numération des colonies sur les membranes, les filtres ont été retirées et les boîtes de Pétri ont été à nouveau incubées pendant 7 jours à 28°C puis comptées à nouveau.

1.8 RECUPERATION DE L'ADN DE PHAGE LAMBDA AJOUTE AUX
10 **SOLS:** L'ADN de phage lambda a été digéré avec HindIII, extrait par un mélange de phénol-chloroforme, précipité puis resuspendu dans de l'eau stérile ultrapure selon des protocoles standard (SAMBROOK et al., 1989).

Des dilutions correspondant respectivement à 0, 2,5, 5, 7,5, 10
15 et 15 µg d'ADN/g de poids sec de sol ont été préparées dans des volumes de 60 µl. Ces dilutions d'ADN ont été ajoutées à des lots de 5g de sol sec qui ont été subséquemment vigoureusement mélangés par vortex pendant 5 minutes avant broyage.

L'ADN de phage lambda a aussi été ajouté à un sol avant
20 broyage à des concentrations correspondant à 0, 10 et 15 µg d'ADN/g de poids sec du sol.

Après broyage, le tampon d'extraction est ajouté et l'ADN est extrait selon le protocole 2 (voir ci-dessus).

25 **1.9 SATURATION DES SITES D'ADSORPTION AVEC DE L'ARN:** Afin de déterminer si la saturation des sites d'adsorption d'acides nucléiques des colloïdes du sol pouvait augmenter le taux de récupération de l'ADN, le terreau sablonneux (sol n°4) et le sol argileux (sol n°5) ont été incubés avec une solution d'ARN avant tout autre traitement.

30 De l'ARN commercial de *Saccharomyces cerevisiae* (BOHRINGER MANNHEIM, MEYLAN, France) a été dilué dans du tampon phosphate (pH 7,1) et ajouté aux échantillons de sol sec et tamisés (2 ml/g de sol) à des concentrations finales de 20, 50 et 100 mg d'ARN/g de poids sec du sol.

Les tubes contenant les suspensions de sol ont été agités par rotation pendant deux heures à température ambiante. Après centrifugation, les culots de sol ont été séchés au four (50°C) pendant la nuit. L'ADN de phage lambda a ensuite été ajouté aux sols (0, 20 ou 50 µg/g de poids sec du sol) afin de simuler le sort de l'ADN libéré après lyse cellulaire.

L'ADN a été extrait selon le protocole n°2. Il a été déterminé par la suite qu'un effet identique de l'addition d'ARN sur la récupération d'ADN pouvait être atteint en ajoutant l'ARN directement au tampon d'extraction.

Cette procédure simplifiée a été utilisée pour le sol argileux n°5 dans les expériences dans lesquelles les micro-organismes ont été inoculés dans les sols.

L'ARN a ensuite été ajouté à une concentration correspondant à 50 mg d'ARN/g de poids sec du sol.

1.10 DETERMINATION QUALITATIVE ET QUANTITATIVE DE L'EFFICACITE DES PROTOCOLES D'EXTRACTION: La qualité de l'ADN (absence de dégradation) a été estimée sur la base de la taille des fragments d'ADN ou de la position relative des bandes de migration d'ADN après électrophorèse d'une fraction aliquote d'une solution d'ADN sur un gel d'agarose à 0,8%.

L'intensité de fluorescence a permis une estimation semi-quantitative des rendements d'extraction.

Une autre fraction aliquote a été utilisée pour des déterminations quantitatives de la teneur en ADN par hybridation (Dot Blot) et analyse au phospho-Imager. Le protocole d'hybridation sur tache a été décrit par SIMONET et al. (1990).

Les membranes d'hybridation (GeneScreen plus, Life Science Products, Boston, Etats-Unis d'Amérique) ont été préhybridées pendant au moins 2 heures dans 20 ml d'une solution contenant 6 ml de 20 x SSC, 1 ml de solution de DENHARDT's, 1 ml de SDS à 10% et 5 mg d'ADN de sperme de saumon.

L'hybridation a été réalisée pendant une nuit dans la même solution en présence d'une sonde marquée préalablement à deux

lavages des membranes dans un tampon SSC 2 x pendant 5 minutes à température ambiante, puis un troisième lavage dans du tampon SSC 2 x, SDS 0,1% et un quatrième lavage dans du tampon SSC 1 x, SDS 0,1% pendant 30 minutes à la température d'hybridation.

5 Les signaux d'hybridation ont été quantifiés avec un système d'imagerie radioanalytique BIORAD (Molecular Analyst Software, BIORAD, Ivry S/Seine, France).

Afin de quantifier la quantité totale d'ADN dérivée de la microflore indigène, les différents sols ont été extraits selon les
10 protocoles n°1 à 5. L'ADN non amplifié a été appliqué sur les membranes de Dot Blot et hybridé en utilisant la sonde universelle FGPS431 (tableau 2).

Cette sonde, qui hybride aux positions 1392-1406 du gène de l'ADNr 16S de *E.coli* (Amann et al. (1995)) a été marquée à ses
15 extrémités avec un $\text{ATP}\alpha^{32}\text{P}$ en utilisant une polynucléotide kinase T4(BOEHRINGER MANNHEIM, Melan, France).

Une courbe de calibration a été préparée à partir de l'ADN de *E.coli* DH5 α . La conversion des calculs aux bactéries du sol a nécessité une simplification, partant de l'hypothèse que le nombre de copies
20 moyen (rrn) est de 7, comme pour *E.coli*.

L'ADN de phage lambda digéré par HindIII a été utilisé pour quantifier la récupération de l'ADN extracellulaire. Des extraits non amplifiés à partir de sols, auxquels de l'ADN de phage lambda avait été ajouté, ont été hybridés avec de l'ADN de phage lambda digéré par
25 HindIII marqué au hasard en utilisant le fragment Klenow (Boehringer Mannheim, Melan, France).

Les quantités d'ADN ont été calculées par interpolation à partir d'une courbe de calibration préparée avec l'ADN purifié.

La quantité totale d'ADN extrait à partir des sols n°1, 2, 3, 4 et 6
30 selon le protocole n°2 (broyage) a également été quantifiée de manière colorimétrique selon la technique décrite par RICHARD (1974).

Brièvement, de l'ADN a été mélangé avec du HClO_4 concentré (la concentration finale de HClO_4 était de 1,5 N). On a mélangé 2,5 volumes de cette solution avec 1,5 volumes de DPA (diphénylamine,
35 Sigma-Aldrich, France) et laissé incuber le mélange à la température

ambiante pendant 18 heures, préalablement à la détermination de la DO à 600 nm. Les extraits d'ADN du sol ont été quantifiés par rapport à une courbe standard réalisée par l'ADN extrait à partir de *E.coli* DH5 α selon les protocoles standards (SAMBROOK et al., (1989)).

5

1.11 DEVELOPPEMENT D'UNE TECHNIQUE DE QUANTIFICATION D'ADN EN UTILISANT L'AMPLIFICATION PCR ET L'HYBRIDATION:

Pour les amplifications par PCR, de l'ADN polymérase Taq (Appligene Oncor, France) a été utilisé selon les instructions du fabricant.

10 Le programme PCR utilisé pour toutes les amplifications est le suivant: dénaturation initiale pendant 3 minutes à 95°C, puis 35 cycles consistant en 1 minute à 95°C, 1 minute à 55°C et 1 minute à 72°C, suivie par une extension finale à 72°C pendant 3 minutes.

15 L'ADN isolé et purifié à partir de *Streptosporangium fragile* a été utilisé comme témoin à des concentrations allant de 100 fg à 100 ng.

Afin d'amplifier spécifiquement l'ADN de ce genre bactérien, on a choisi les amorces FGPS122 et FGPS350 (tableau 2), complémentaires à une partie de l'ADNr 16S, après alignement des séquences d'ADNr 16S d'actinomycètes. Leur spécificité a été testée
20 sur une collection de souches d'actinomycètes (*Streptomycès*, *Streptosporangium* et d'autres genres fortement apparentés).

Les produits de PCR ont été hybridés avec la sonde oligonucléotidique FGPS643 (tableau 2). Afin de simuler le niveau de pureté obtenu en routine avec de l'ADN extrait à partir du sol, des
25 témoins d'ADN pur de *S. fragile* ont été mélangés avec les extraits de sol obtenus après des traitements selon les protocoles de lyse 4b et 5b puis purifiés selon le protocole D.

Avant utilisation, les extraits de sol ont été traités avec de la DNase (une unité de DNase/ml, GIBCO BRL) pendant 30 minutes à
30 température ambiante. La DNase a ensuite été inactivée par chauffage à 65°C pendant 10 minutes. Une vérification de l'inactivation a été réalisée par PCR. Les concentrations d'acides humiques ont été mesurées par spectrophotométrie (DO_{280nm}) contre une courbe standard d'acides humiques commerciaux (Sigma).

Des solutions de sol traitées à la Dnase non diluées, diluées 10x et diluées x 100 ont été mélangées de 100. fg à 100ng d'ADN de *S. fragile* avant l'amplification par PCR. Dans une autre série d'expériences, les concentrations croissantes d'ADN de *Streptomyces*
5 *hygroscopicus* de (100 pg à 1 µg) ont été ajoutées à l'ADN de *S. fragile* afin de simuler la présence d'ADN non-cible et son influence sur le procédé PCR.

1.12 PURIFICATION DES EXTRAITS D'ADN BRUT: Quatre méthodes
10 de purification d'ADN ont été comparées. L'ADN a été extrait à partir de 1g (poids sec de sol selon le protocole 4a et remis en suspension dans 100 µl de tampon TE8 (50 mM Tris, 20 mM (EDTA, pH 8,0).

Protocole A

15

Elution à travers deux colonnes successives Elutip d (SCHLEICHER et SCHUELL, Dassel, Allemagne) (PICARD et al., (1992)).

20

Protocole B:

Elution à travers une colonne SEPHACRYL S200 (Pharmacia Biotech, Uppsala, Suède) suivie d'une élution à travers une colonne Elutip d (NESME et al. (1995)).

25

Protocole C:

Séparation à l'aide d'un système aqueux à deux phases avec 17,9% (poids/poids) de PEG 8000 (Merck, Darmstadt, Allemagne) et
30 14,3% (poids/poids) de (NH₄)₂SO₄(ZASLAVSKY,(1995)).

Après un mélange vigoureux au vortex, les deux phases ont été laissées à température ambiante pour leur séparation.

1 ml de chacune des phases a été transféré dans un autre tube, mélangé avec 100µl de l'échantillon et laissé à 4°C pendant une
35 nuit pour permettre la séparation.

La phase inférieure a été dialysée pendant une heure à travers une membrane Millipore en présence d'un excès d'un tampon TE 7,5 (10 mM Tris, 1 mM EDTA à pH 7,5 et 1 M Mg Cl₂) afin d'éliminer les sels en excès.

5

Protocole D:

Elution à travers une colonne de type Microspin Sephacryl S400 HR (Pharmacia Biotech, Uppsala, Suède) , suivie d'une élution à
10 travers une colonne de type Elutip d.

Chaque protocole est terminé par une étape de précipitation à l'éthanol, et l'ADN est remis en suspension dans 10 µl de tampon TE 7,5. L'efficacité des protocoles de purification a été vérifiée par amplification PCR de fractions aliquotes non diluées des solutions d'ADN
15 et de fractions aliquotes diluées 10 et x 100 fois, en utilisant des protocoles standard (voir ci-dessous).

1.13 RECUPERATION DE L'ADN A PARTIR DE MICROORGANISMES INNOCULES:

20 Les cellules, spores et hyphae ont été lavées deux fois et dénombrées par comptage sur plaque ou comptage microscopique direct. Des lots de 5g de sol sec et tamisé (sols n°2, 3 et 5) ont été inoculés avec 100 µl d'une suspension de spores et d'hyphae de *S. lividans* à des concentrations correspondant à 0,10³, 10⁵, 10⁷ et 10⁹ spores/g de poids
25 sec de sol, ou avec des cellules végétatives de *B.anthraxis* à des concentrations correspondant à 0,10⁷ et 10⁹ cellules par gramme de poids sec du sol.

Les quantités de hyphae de *S. lividans* ont été calculées sur la base du nombre de spores desquelles elles sont originaires. Après
30 addition des suspensions bactériennes, les échantillons de sol sont mélangés vigoureusement par vortex pendant 5 minutes avant broyage. L'ADN est extrait selon le protocole n°6 (voir ci-dessous).

L'amplification PCR suivie d'une hybridation sur tache (Dot Blot) et imagerie par phosphorescence (phospho-imaging) a été utilisée afin

de quantifier les quantités d'ADN récupérées à partir des cellules, des spores et du mycélium bactérien inoculé dans les sols.

L'extraction d'ADN a été réalisée selon le protocole de lyse n°6. L'amplification PCR et l'hybridation ont été réalisées comme décrit ci-dessus. Les amorces et les sondes sont ciblées sur des régions chromosomiques localisées en dehors de la région 16S, et sont hautement spécifiques des organismes respectifs, de manière à éviter des signaux de bruit de fond.

Pour les sols ensemencés avec *B. anthracis*, les amorces R499 et R500 ont été utilisées (Patra et al. (1996)) et les produits d'amplification ont été hybridés avec la sonde oligonucléotidique C501 (tableau 2).

Pour les sols ensemencés avec *S. lividans*, les réactions PCR ont été réalisées en utilisant les amorces FGPS516 et FGPS517, et les produits d'amplification ont été hybridés avec la sonde oligonucléotidique FGPS518 (tableau 2).

La région amplifiée est une partie de la cassette construite spécifiquement pour obtenir la souche OS48.3 (CLERC-BARDIN et al., non publié).

Les comptes de calibration ont été dans tous les cas obtenus en utilisant l'ADN purifié de l'organisme cible.

2. RESULTATS

2.1 CHOIX DU TAMPON D'EXTRACTION:

20 sols différents ont été utilisés afin de déterminer le pH optimal du tampon d'extraction d'ADN. Pour tous les sols, le rendement en ADN augmente avec les pH croissants du tampon. Le rendement pour chaque pH (+/- sd), calculé comme le pourcentage de la valeur la plus haute pour chacun des sols, est le suivant: pH 6,0 : 31 +/- 13; pH 7,0: 43 +/- 16; pH 8,0: 60 +/- 14; pH 9,0: 82 +/- 12; pH 10,0: 98 +/- 3.

Pour 16 des 20 sols, le rendement le plus élevé a été obtenu à pH 10,0, alors que pour les quatre autres sols le plus haut rendement a

été obtenu à pH 9,0. Cependant, à pH 10,0, des quantités plus grandes de matériel humique ont été libérées, comparées à pH 9,0 (résultats non présentés). En conséquence, le pH 9,0 a été choisi pour toutes les expériences présentées ci-dessous.

5

2.2 EFFICACITE DES PROTOCOLES D'EXTRACTION D'ADN:

L'ADN total des organismes indigènes du sol a été extrait et quantifié de manière à évaluer l'efficacité de nombreux protocoles de lyse cellulaire *in situ*. Des échantillons des sols 1-6 (tableau 1) ont été traités selon les protocoles n°1 à 5 décrits dans la section Matériel et Méthodes (figure 1).

Après l'extraction d'ADN, les suspensions de sols ont été précipitées avec de l'isopropanol, et des fractions aliquotes des culots remis en suspension ont été analysées par électrophorèse sur gel, dans une première étape, afin d'estimer la qualité et la quantité de l'ADN libéré.

Cependant, la couleur de l'extrait d'ADN devenait de plus en plus sombre au fur et à mesure du nombre croissant d'étapes de lyse, du fait de la co-extraction de composés, tels que les acides humiques, avec l'ADN.

Certains de ces extraits bruts de couleur sombre ne migrent pas de la manière attendue dans les gels d'agarose.

En conséquence, les solutions d'ADN brut ont été purifiées (protocole B) avant quantification. Les électrophorèses sur gel des solutions purifiées obtenues après les différents traitements de lyse sont exemplifiées sur le sol n°3 (figure 2).

Une comparaison visuelle au rayonnement ultra-violet des intensités de l'ADN coloré a permis une estimation semi-quantitative de l'efficacité des traitements. De plus, la présence de profils de migration de tailles multiples de fragments (bandes discrètes) d'ADN et la disparition des fragments longs indique qu'une dégradation de l'ADN a eu lieu.

Aucun ADN n'a pu être extrait du sol argileux n°5.

Une quantification plus précise de l'ADN de tous les sols, extrait selon les protocoles n°1 à 5, a été réalisée par hybridation sur tache (Dot Blot) sans étape d'amplification PCR préalable et en utilisant une sonde oligonucléotidique complémentaire d'une séquence hautement conservée de la région d'ADNr 16S (sonde FGPS 431, tableau 2).

L'ADN a été détecté dans les extraits de tous les sols après chacune des différentes étapes de lyse, à l'exception du sol argileux n°5.

Les résultats concordent avec les estimations réalisées après gel d'électrophorèse.

Afin de comparer avec une méthode indépendante pour la quantification, l'ADN extrait selon le protocole n°2 (tous les sols sauf le sol n°5) a été également quantifié en utilisant une méthode colorimétrique de détection de l'ADN (RICHARD, 1974).

On a trouvé une bonne corrélation ($r = 0,88$) entre l'ADN quantifié en utilisant cette technique colorimétrique et les résultats obtenus par hybridation de type Dot Blot/radio-imagerie, confirmant l'hypothèse selon laquelle le nombre de copies moyen des bactéries du sol (rrn) est de 7.

L'hybridation (Dot Blot) a montré que les quantités d'ADN extracellulaires, comme déterminé par extraction sans traitement de lyse (protocole n°1), allait de 4 µg/g pour le sol acide (n°6) à 36 µg/g pour le sol n°3 (tableau 3).

Le broyage du sol (protocole n°2) a augmenté les quantités d'ADN extrait à partir de tous les sols (p.ex. 26 µg/g de sol) pour le sol n°6 et 59 µg/g de sol (pour le sol n°3) (tableau 3; figure 2).

Pour les deux traitements de broyage (voir la section Matériel et Méthodes) la migration discrète d'ADN a été détectée sur les gels d'agarose, indiquant que les molécules d'ADN ont été partiellement dégradées (figure 2).

La taille des fragments d'ADN est comprise entre 20 et 0,2 kb. L'intensité de bande des fragments les plus petits est très faible, indiquant que la majeure partie des fragments ont une taille bien supérieure à 1 kb.

Le protocole n°3 comprend une étape d'homogénéisation dans un dispositif mixeur de type Ultraturax après l'addition du tampon

d'extraction aux échantillons de sol. Cette étape conduit à une augmentation des quantités d'ADN extrait, comme déterminé par hybridation sur tache (Dot Blot) pour deux des sols (le terreau sablonneux n°3 et le sol acide n°6), alors que les deux sols riches en matière organique (sols n°1 et n°2) ont conduit à l'obtention de quantités plus faibles d'ADN.

Les protocoles n°4a et n°4b ont permis d'évaluer l'influence de deux types de sonication sur les rendements en ADN à partir de sols préalablement broyés et homogénéisés .

La sonication n'a pas eu d'effet positif sur le rendement en ADN, comparé au protocole n°3, excepté pour le sol n°6. Toutefois, l'efficacité de lyse des deux types de sonicateur diffèrent. Pour les sols n°2, 3 et 4, les quantités d'ADN extraits les plus grandes ont été obtenues en utilisant la micropointe de titane (tableau 3; figure 2), alors que pour les sols n°1 et n°6, le rendement en ADN était supérieur en utilisant le dispositif Cup Horn.

Des résultats contradictoires ont été également obtenus lorsque l'on a ajouté une étape de lyse enzymatique/chimique (protocoles n°5a et 5b) après l'étape de sonication: dans certains cas, les quantités d'ADN extraites ont été plus grandes que celles récupérées selon les protocoles n°4a et 4b, alors que dans d'autres cas les rendements étaient moindres (tableau 3).

2.3 COMPTAGE DIRECT DES MICRO-ORGANISMES:

Des comptes au microscope du nombre total de cellules bactériennes après coloration à l'acridine orange ont été réalisés pour tous les sols, avant et après broyage.

Avant broyage, le nombre de bactéries par gramme de poids sec du sol allait de $1,4 \times 10^9$ (+/- 0,4) dans le sol tropical n°5 à 10×10^9 (+/- 0,7) dans le sol provenant de la Côte Saint-André (sol n°3) (tableau 1).

Après broyage, les nombres de cellules ont été respectivement de 45, 74, 75, 54, 34 et 75% des valeurs initiales pour les sols n°1 à 6.

2.4 NUMERATION DES ACTINOMYCETES CULTIVABLES APPARTENANT A DIFFERENTS GENRES:

Une modification dans les populations d'actinomycètes dans le sol n°3 a été remarquée après les différents traitements de lyse (figure 3).

Par exemple, les colonies de *Streptomyces* sp. dominaient la flore viable d'actinomycètes lorsqu'aucun traitement de lyse n'est appliqué (protocole n°1), et représentaient 65% du nombre total de colonies identifiées. Après broyage, le pourcentage de colonies de *Streptomyces* a diminué pour atteindre 51%, alors que la proportion de colonies appartenant au genre *Micromonospora* a augmenté de 14% à 41%.

La lyse chimique/enzymatique (protocoles 5a et 5b) est apparue comme particulièrement efficace pour la lyse des streptomycètes. Lorsque tous les traitements de lyse ont été appliqués, y compris une lyse chimique/enzymatique (protocoles 5a et 5b), la microflore d'actinomycètes, qui comprenait encore plus de 10^6 CFU/g de sol, était dominée par les espèces appartenant au genre *Micromonospora*, alors qu'aucune ou très peu de colonies de *Streptomyces* ont été récupérées.

Les organismes appartenant aux genres tels que *Streptosporangium*, *Actinomadura*, *Microbispora*, *Dactylosporangium* et *Actinoplanes* sont apparus sur les plaques en faible nombre (2-8% du nombre total de colonies identifiées) après broyage, homogénéisation avec le dispositif Ultraturrax, et sonication, mais étaient généralement absents lorsque ces traitements étaient combinés avec une lyse chimique/enzymatique.

Le nombre total de bactéries cultivables restant après chaque traitement de lyse (protocoles 2 à 5) a été aussi recherché pour le sol n°4. Les résultats indiquent que le nombre de bactéries cultivables ne décroît pas avec l'intensité des traitements de lyse (environ 2×10^6 CFU/g de sol dans tous les cas, et également lorsqu'un traitement n'est appliqué, tel que selon le protocole n°1).

L'obtention de ces faibles valeurs de CFU est probablement due au fait que du sol sec a été utilisé et que seules les bactéries les

plus résistantes se sont multipliées sur les plaques. Le nombre d'actinomycètes formant colonies était généralement plus grand que celui des CFU total (toutes les bactéries) du fait qu'une étape de germination de spores, comprise dans le protocole de détection des actinomycètes, manquait lors du contrôle des bactéries totales.

2.5 RECUPERATION DE L'ADN DU PHAGE LAMBDA AJOUTE:

Le but de ces expériences était d'estimer de quelle manière des traitements de lyse successifs pouvaient affecter la récupération d'ADN nu, et si ces traitements successifs de lyse contribuaient à sa dégradation.

L'ADN pouvait être soit une fraction d'ADN extracellulaire libérée à partir d'organismes déjà morts, qui peuvent persister dans le sol pendant des mois (WARD et al., 1990), soit de l'ADN libéré à partir d'organismes lysés facilement pendant les premières étapes du traitement. Afin de simuler cette situation, de l'ADN de phage lambda digéré par HindIII a été ajouté, à diverses concentrations, aux sols avant et après broyage. En plus du broyage, une combinaison des autres traitements de lyse a été testée, y compris la sonication (dispositif Cup Horn, voir protocole n°4b) et des chocs thermiques (voir la section Matériel et Méthodes).

Après extraction, des fractions aliquotes qui devraient théoriquement contenir de 25 à 150 ng d'ADN de phage lambda ont été analysées par électrophorèse sur gel. Aucun fragment d'ADN spécifique du phage lambda n'a pu être observé lorsque l'ADN a été inoculé dans les échantillons de sol préalablement au broyage, indépendamment de la dose ou du type de sol.

Lorsque l'ADN a été ajouté après broyage, et extrait sans étape de traitement de lyse additionnelle, les profils spécifiques d'ADN de phage lambda ont été détectés dans les extraits de quatre des cinq sols testés.

Dans tous ces cas, une relation directe de cause à effet a été obtenue entre la quantité d'ADN ajoutée et l'intensité des signaux sur les gels d'agarose. Les intensités des signaux étaient, cependant,

inférieures aux intensités de signaux attendues si on les compare à celles des standards moléculaires.

De plus, la bande à 23 kb était absente dans plusieurs cas, indiquant que les longs fragments étaient préférentiellement adsorbés aux particules du sol, ou étaient plus sensibles à la dégradation, comparés aux fragments courts.

Aucune bande n'a été détectée dans les échantillons de sol tropical n°5 qui est caractérisé par une très haute teneur en argile (tableau 1).

Pour une quantification plus précise, la récupération d'ADN a été déterminée sur un dispositif d'imagerie par phosphorescence (phospho-imager) après hybridation en tache (Dot Blot). Selon cette technique, l'ADN a été détecté dans tous les échantillons, y compris ceux qui avaient été inoculés avant broyage, à l'exception du sol n°5 dans lequel aucun ADN n'a pu être détecté.

Dans tous les autres sols, la quantité d'ADN extrait augmente avec l'augmentation de taille de l'inoculum (figures 4a-d).

Cependant, les récupérations d'ADN de phage lambda étaient faibles. Lorsque le broyage était le seul traitement de lyse appliqué, les récupérations étaient comprises entre 0,6 et 5,9% de l'ADN ajouté lorsque celui-ci était ajouté avant broyage, et de 3,6 à 24% de l'ADN ajouté lorsque ce dernier était ajouté après broyage. Les plus hauts niveaux de récupération ont été obtenus à partir du sol n°2.

L'électrophorèse sur gel de fractions aliquotes d'échantillons traités par choc thermique et sonication n'a permis d'observer des bandes d'ADN dans aucun des échantillons, y compris l'essai dans lequel l'ADN avait été ajouté après broyage. Les expériences d'hybridation en tache (Dot Blot) ont confirmé ces résultats.

Les signaux d'hybridation obtenus à partir de suspensions de sol qui ont été traitées par chocs thermiques et sonication ont été, tout au plus, faibles.

L'échantillon présentant la plus forte quantité d'ADN (15 µg d'ADN/g de poids sec du sol) était le seul pour lequel le signal obtenu était sensiblement différent du niveau du bruit de fond.

Aucune différence, (ou de faibles différences) n'a été observée entre les échantillons traités par choc thermique et ceux traités par chocs thermiques et sonication, indiquant que les chocs thermiques ont un effet préjudiciable sur l'ADN. Les récupérations les meilleures ont été observées pour le sol n°2, qui a la plus forte teneur en matière organique (tableau 1), alors qu'aucun ADN n'a été récupéré à partir du sol argileux n°5.

Des expériences additionnelles ont été réalisées avec des échantillons non broyés de sols n°4 et n°5, qui ont été ensemencés avec 20 et 50 µg d'ADN de phage lambda par gramme de sol.

Les échantillons ont été extraits immédiatement ou après une période d'incubation d'une heure à 28°C, puis les extraits d'ADN ont été purifiés et analysés par électrophorèse sur gel.

L'incubation du sol n°4 pendant une heure après l'inoculation n'a pas conduit à des profils qualitativement ou quantitativement différents de ceux obtenus sans incubation ou de ceux observés antérieurement lorsque l'ADN avait ajouté après broyage.

Ces résultats indiquent que la dégradation enzymatique par les nucléases du sol ne seraient pas impliquée dans le faible taux de récupération d'ADN. De plus, l'absence d'étape de broyage ne permet pas une augmentation de la récupération de l'ADN à partir du sol n°5, indiquant que les modifications de structure du sol dues au broyage n'augmentent pas significativement l'adsorption des acides nucléiques sur les colloïdes.

2.6 SATURATION DES SITES D'ADSORPTION AVEC L'ARN:

La plupart des profils obtenus sur les gels d'agarose ne diffèrent pas significativement des profils précédents dans lesquels le traitement d'ARN n'a pas été effectué.

Par exemple, aucune bande n'a été détectée à partir du sol riche en argile n°5, indépendamment des concentrations d'ARN et des concentrations d'ADN de phage lambda utilisées.

De plus, les bandes spécifiques d'ADN de phage lambda digérées par HindIII restaient indétectables dans le terreau sablonneux traité par l'ARN (sol n°4) lorsque l'ARN est ajouté avant le broyage.

5 L'intensité des bandes obtenues à partir d'échantillons ensemencés avec l'ADN après broyage augmente avec la concentration d'ARN, indiquant que le traitement pourrait avoir un effet positif.

Cependant, les résultats après hybridation et analyse par imagerie à phosphorescence n'ont pas confirmé les résultats de l'électrophorèse. Par exemple, l'effet positif du traitement d'ARN sur la
10 récupération d'ADN à partir du terreau argileux, lorsque l'ADN a été ajouté après broyage, n'apparaît pas clairement.

D'un autre côté, un effet positif de l'ARN a été trouvé pour le sol riche en argile (n°5) lorsque l'ADN a été ajouté après broyage.

Bien que les signaux d'hybridation pour les échantillons
15 contrôle ne diffèrent pas des niveaux de bruit de fond, des quantités significatives d'ADN ont été libérées à partir des échantillons traités par l'ARN, et les signaux ont augmenté avec la quantité d'ADN ajoutée ainsi qu'avec la concentration d'ARN.

Cependant, même pour la plus forte concentration d'ARN (100
20 mg/g de poids de sol sec) le taux de récupération n'a jamais dépassé 3%.

2.7 PURIFICATION DES EXTRAITS BRUTS D'ADN:

25 Des quatre protocoles testés, la meilleure amplification des extraits d'ADN non dilués (1 µl d'extrait dans 50 µl de mélange PCR) a été observée après l'élution à travers des colonnes de type Microspin S400 suivie d'une élution à travers une colonne de type Elutip d, comme le montre l'électrophorèse sur gel des produits PCR.

30 L'ADN purifié par le système aqueux double phase (protocole C) a donné des quantités plus faibles de produits PCR après amplification à partir d'extrait d'ADN non dilué.

Aucun produit d'amplification n'a pu être obtenu à partir des extraits non dilués après amplification à la suite de la mise en oeuvre
35 des protocoles A ou B. En conséquence, le protocole B (voir section

Matériels et Méthodes) a été utilisé pour toutes les expériences dans lesquelles les amplifications PCR et/ou les hybridations sur tâche (Dot Blot) ont été réalisées.

5 2.8 QUANTIFICATION PAR PCR ET HYBRIDATION:

La première étape était de déterminer si les quantités de produit PCR étaient proportionnelles au nombre de molécules d'ADN cibles initialement présentes dans le tube réactionnel. De l'ADN de
10 *Streptosporangium fragile* a été utilisé comme cible (voir section Matériels et Méthodes).

Les amorces utilisées ont été les amorces FGPS122 et FGPS350 (tableau 2). L'électrophorèse sur gel des produits PCR a montré que l'intensité de bande augmente avec l'accroissement de la
15 concentration des cibles. Les produits PCR ont été hybridés avec la sonde oligonucléotidique FGPS643 (tableau 2), et les signaux ont été quantifiés par imagerie par phosphorescence (phospho-imaging).

On a trouvé une bonne corrélation ($r^2 = 0,98$) entre le log[nombre de cibles] et le log[intensité du signal d'hybridation].

20 On a ensuite recherché si l'efficacité de l'amplification PCR était affectée par les acides humiques et l'ADN non cible. Lorsqu'on l'analyse par électrophorèse sur gel, l'intensité accrue des bandes des produits PCR, correspondant aux différentes quantités d'ADN cible, était conservée lorsque l'amplification était réalisée avec des solutions d'ADN
25 auxquelles on avait ajouté des extraits de sol traités à la DNase, contenant des acides humiques à des concentrations allant jusqu'à 8ng dans le mélange PCR d'un volume de 50 µl.

Avec 20 ng d'acide humique dans le mélange PCR, les bandes correspondant aux faibles niveaux d'ADN cible ont disparu, et à des
30 concentrations d'acide humique de 80 ng et à des concentrations supérieures, aucune bande n'était visible.

Les quantités variées d'ADN cible de *S.fragile* ont permis de fournir les quantités attendues de produit PCR lorsque, avant amplification, l'ADN de *S. fragile* a été mélangé avec de l'ADN de
35 *Streptomyces hygroscopicus* et ajouté au mélange PCR de 50 µl dans

une gamme de 100 pg à 1µg afin de simuler l'ADN non-cible libéré à partir de la microflore du sol.

2.9 QUANTIFICATION DES ACTINOMYCETES INDIGENE DU SOL APRES DIFFERENTS TRAITEMENTS DE LYE:

On a appliqué le protocole de purification D suivi d'une amplification par PCR comme décrit ci-dessus afin de quantifier les actinomycètes appartenant au genre *Streptosporangium* dans le sol n°3 après extraction conformément aux protocoles n°1, 2, 3, 5a et 5b (figure 5).

Après broyage, (protocole n°2) la quantité d'ADN cible provenant de cet actinomycète a été estimée par hybridation (Dot Blot) et radio-imagerie comme étant de 2,5 +/- 1,3 ng /g de poids de sol sec.

Si l'on postule que le contenu en ADN est de 10 fg par cellule, comme pour *Streptomyces* (Gladek et al. 1984), cette valeur correspond à approximativement $2,5 \times 10^5$ génomes. Des valeurs similaires ont été obtenues après les autres traitements de lyse (respectivement 2,6 +/-1,1 et 1,8 +/- 1,3 ng d'ADN/g de sol sec en utilisant respectivement les protocoles 3 et 4b).

2.10 EFFICACITE DE LA RECUPARATION D'ADN A PARTIR DE SOLS PREALABLEMENT INOCULES AVEC DES BACTERIES:

Trois sols (n°2, 3 et 5) ont été inoculés avec des spores ou des hyphae de *Streptomyces lividans* à différentes concentrations (voir section Matériel et Méthodes). Les quantités de mycélium ajoutées au sol (figure 6b) correspondent au nombre de spores inoculées dans le milieu de germination. Approximativement 50% de ces spores ont germé. Le nombre exact de cellules dans les hyphae des spores germinées n'a pas été déterminé. En conséquence, les quantités de spores et de mycélium ensemencées dans les sols ne sont pas directement comparables.

Pour chaque échantillon de sol, le protocole d'extraction n°6, la méthode de purification D, et l'amplification PCR combinée avec l'hybridation sur tache (Dot Blot) et l'imagerie par phosphorescence (phospho-imaging) ont été utilisés pour dénombrer les ADNs cibles

spécifiques qui avaient été libérés. L'ADN extrait peut être clairement distingué du bruit de fond seulement lorsque le nombre de spores ajoutées dépasse 10^5 pour les sols n°3 et n°5 et 10^7 pour le sol n°2 (figure 6a).

5 Lorsque le mycélium est ajouté, l'ADN extrait peut être détecté au-delà d'une quantité correspondant à 10^3 spores/g de sol pour les sols n°2 et n°3, et au-delà de 10^7 spores/g pour le sol n°5 (figure b).

Au-dessus du niveau de détection, le signal d'hybridation augmente avec des quantités croissantes des cellules inoculées.

10 Pour l'inoculum de spores, une augmentation de 100 fois dans le nombre de cellulesensemencées conduit à une augmentation de presque 100 du rendement d'ADN. Cette augmentation est clairement inférieure lorsque les hyphae sont inoculées, particulièrement dans les sols n°2 et n°3 (figure 6).

15 Au contraire, les résultats obtenus lorsque l'ADN de phage lambda a été utilisé comme inoculum, l'ADN a également été récupéré à partir du sol riche en argile (n°5) lorsque les cellules bactériennes ont été utilisées comme inoculum. Cependant, pour ce dernier aussi, le traitement par l'ARN a augmenté la récupération d'ADN de
20 *Streptomyces* à partir de ce sol à la fois pour les spores et le mycélium (figure 6).

Le fait d'ensemencer des sols avec des cellules végétatives de *Bacillus anthracis* a fourni des taux de récupération similaires à ceux obtenus pour *Streptomyces*.

25 De plus, les taux de récupération d'ADN à partir du sol n°5 ont augmenté après traitement par l'ARN également pour cet inoculum.

**Exemple 2 : Construction d'une banque d'ADN de faible poids moléculaire (<10 kb) à partir d'un sol contaminé par du lindane :
30 clonage et expression du gène *linA***

Cet exemple décrit la construction d'une librairie d'ADN du sol dans *E. coli*. Il permet de démontrer le clonage et l'expression de gènes de petite taille issus d'une microflore non cultivable .

Le lindane est un pesticide organochloré, récalcitrant à la dégradation et persistant dans l'environnement. En aérobie, sa biodégradation est catalysée par une déhydrochlorinase, codée par le gène *linA*, permettant de transformer le lindane en 1,2,4-trichlorobenzène. Le gène *linA* n'a été identifié que parmi deux souches isolées du sol : *Sphingomonas paucimobilis*, isolé au Japon (Seeno et Wada 1989, Imai *et al* 1991, Nagata *et al* 1993) et *Rhodanobacter lindaniclasticus* isolé en France (Thomas *et al* 1996, Nalin *et al* 1999).

Pourtant le potentiel de dégradation du lindane, mis en évidence par dosage des ions chlorures libérés et amplification par PCR du gène *linA* à partir de sols ayant été en contact ou non avec du lindane, semble être répandu plus largement dans l'environnement (Biesiekierska-Galgien, 1997).

1. Extraction directe d'ADN de sol

Les sols secs sont broyés pendant 10 minutes dans un broyeur à force centrifuge Restch équipé 6 billes de tungstène. 10 grammes de sol broyé sont mis en suspension dans 50 ml de tampon TENP pH 9 (Tris 50 mM, EDTA 20 mM, NaCl 100 mM, polyvinylpyrrolidone 1% w/v), et homogénéisés au vortex pendant 10 min.

Après centrifugation de 5 minutes, 4000 g à 4°C, le surnageant est précipité à l'acétate de sodium (3M, pH 5.2) et à l'isopropanol, pour être repris dans du tampon TE stérile (Tris 10 mM, EDTA 1 mM, pH 8.0). L'ADN extrait est ensuite purifié sur colonne de tamisage moléculaire S400 (Pharmacia) et sur colonne échangeuse d'ions Elutip d (Schleicher et Schuell), selon les instructions des fabricants, puis conservé dans du TE.

2. Construction de la banque d'ADN extrait du sol dans le vecteur pBluescript SK-

Le vecteur pBluescript SK- et l'ADN extrait du sol sont chacun digérés par les enzymes *HindIII* et *BamHI* (Roche), à raison de 10 unités d'enzymes pour 1 µg d'ADN (incubation 2 heures à 37°C). Les ADN sont

ensuite ligués par action de la T4 DNA ligase (Roche), une nuit à 15°C, à raison d'une unité d'enzyme pour 300 ng d'ADN (environ 200 ng d'ADN insert et 100 ng de vecteur digéré). Les cellules d'*Escherichia coli* électrocompétentes, ElectroMAX DH10B™ (Gibco BRL) sont transformées par le mélange de ligation (2 µl) par électroporation (25 µF, 200 et 500 Ω, 2,5 kV) (Biorad Gene Pulser).

Après une heure d'incubation dans le milieu LB, les cellules transformées sont diluées de façon à obtenir environ 100 colonies par boîte puis sont étalées sur milieu LB (10 g/l Tryptone, 5 g/l extrait de levure, 5 g/ NaCl) additionné d'Ampiciline (100 mg/l), de γ-HCH (500 mg/l), de X-gal 60 mg/l (5-bromo-4-chloro-3-indolyl-α-D-galactoside), et d'IPTG 40 mg/l (isopropylthio-β-D-galactoside), et incubées une nuit à 37°C. Le γ-hexachlorocyclohexane (Merck-Schuchardt) étant insoluble dans l'eau, une solution à 50 g/l est préparée dans du DMSO (diméthyl sulfoxyde) (Sigma).

Une banque de 10 000 clones a ainsi été obtenue.

3. Clonage et expression du gène *linA*

Le criblage de la banque s'effectue par visualisation d'un halo de dégradation du lindane autour de la colonie (le lindane précipitant dans les milieux de culture). Sur 10 000 clones criblés, 35 présentaient ainsi une activité de dégradation du lindane. La présence du gène *linA* chez ces clones a pu être confirmée par PCR grâce à des amorces spécifiques, décrites par Thomas et al (1996). Des digestions réalisées sur les inserts ainsi que sur les produits d'amplification ont montré des profils identiques entre tous les clones criblés et le témoin de référence, *R. lindaniclasticus*. Les clones portant le gène *linA* présentaient également un insert de même taille (environ 4 kb).

Il a ainsi pu être démontré que l'ADN du sol pouvait être cloné et exprimé chez un hôte hétérologue : *E. coli*, et que des gènes issus d'une microflore difficilement cultivable pouvaient être exprimés. Des banques réalisées à partir de digestion partielle d'ADN extrait du sol par des enzymes de restriction telles que *Sau3AI* sont donc aussi envisageables.

EXEMPLE 3:

Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol, comprenant une étape d'extraction indirecte de l'ADN.

1. MATERIEL ET METHODES.**1.1 Extraction de la fraction bactérienne du sol.**

5g de sol sont dispersés dans 50 ml de NaCl 0.8% stérile, par broyage au Waring Blender pendant 3 x 1 minute, avec refroidissement dans la glace entre chaque broyage. les cellules bactériennes sont alors séparées des particules du sol par centrifugation sur un coussin de densité de Nycodenz (Nycomed Pharma AS, Oslo, Norvège). Dans un tube à centrifugation, 11,6 ml d'une solution de Nycodenz de densité de 1.3 g.ml⁻¹ (8g de Nycodenz suspendu dans 10 ml d'eau stérile) sont placés en dessous de 25 ml de la suspension de sol précédemment obtenue. Après centrifugation à 10.000 g dans un rotor à godets mobiles (rotor TST 28.38, Kontron) pendant 40 minutes à 4°C, l'anneau cellulaire, se situant à l'interphase de la phase aqueuse et de la phase Nycodenz, est prélevé, lavé dans 25 ml d'eau stérile et centrifugé à 10.000 g pendant 20 minutes. Le culot cellulaire est ensuite repris dans une solution Tris 10 mM; EDTA 100 mMn pH 8.0.

Préalablement à la dispersion du sol au Waring Blender, une étape d'enrichissement du sol dans une solution d'extrait de levure peut être incluse afin de permettre notamment la germination des spores bactériennes du sol. 5 g de sol sont alors incubés dans 50 ml d'une solution stérile de NaCl 0.8% - extrait de levure 6%, pendant 30 minutes à 40°C. L'extrait de levure est éliminé par centrifugation à 5000 rpm pendant 10 minutes afin d'éviter la formation de mousse durant le broyage.

1.2 Lyse des cellules bactériennes du sol.

- *Lyse des cellules en milieu liquide et purification sur gradient de chlorure de césium.*

Les cellules sont lysées dans une solution Tris 10 mM, EDTA 100 mM, pH 8.0 contenant 5 mg.ml⁻¹ de lysozyme et 0.5 mg.ml⁻¹ d'achromopeptidase pendant 1 heure à 37°C . Une solution de lauryl sarcosyl (1% final) et de protéinase K (2 mg.ml⁻¹) est ensuite ajoutée et incubée à 37°C pendant 30 minutes. La solution d'ADN est alors purifiée sur un gradient de densité de chlorure de césium par centrifugation à 35 000 rpm pendant 36 heures sur un rotor Kontron 65.13. Le gradient de chlorure de césium employé est un gradient à 1g/ml de CsCl, possédant un indice de réfraction de 1,3860 (Sambrook et al., 1989).

- *Lyse des cellules après inclusion dans un bloc d'agarose.*

Les cellules sont mélangées à un volume égal d'agarose à 1.5% (poids/volume) Seaplaque (Agarose Seaplaque FMC Products. TEBU, Le Perray en Yvelines, France). à bas point de fusion et coulées dans un bloc de 100 µl. Les blocs sont ensuite incubés dans une solution de lyse : EDTA 250 mM, saccharose 10.3%, lysozyme 5 mg.ml⁻¹ et achromopeptidase 0.5 mg.ml⁻¹ à 37°C pendant 3 heures. Les blocs sont alors lavés dans une solution de Tris 10 mM - EDTA 500 mM et incubés une nuit à 37°C dans de l'EDTA 500 mM contenant 1 mg.ml⁻¹ de protéinase K et du lauryl sarcosyl 1%. Après plusieurs lavages dans du Tris-EDTA, les blocs sont conservés dans de l'EDTA 500 mM.

La qualité des ADN ainsi extraits est contrôlée par électrophorèse en champs pulsés.

La quantité d'ADN extrait a été évaluée sur gel d'électrophorèse par rapport à une gamme étalon d'ADN de thymus de veau.

1.3 Caractérisation moléculaire de l'ADN extrait du sol.

Les ADN extraits du sol sont caractérisés par hybridation PCR, méthode qui consiste à amplifier dans un premier temps les ADNs à l'aide d'amorces situées sur des régions universellement conservées du gène de l'ARNr16S, puis à hybrider les ADNs amplifiés avec différentes sondes oligonucléotidiques de spécificité connue (tableau 4), dans le but

de quantifier l'intensité du signal d'hybridation par rapport à une gamme étalon externe d'ADN génomique.

Les ADN extraits du sol ainsi que les ADN génomiques extraits de cultures pures sont amplifiés avec les amorces FGPS 612-669 (tableau 1) dans les conditions standard d'amplification par PCR. Les produits d'amplification sont ensuite dénaturés par un volume égal de NaOH 1N, déposés sur une membrane de Nylon (GeneScreen Plus, Life Science Products) et hybridés avec une sonde oligonucléotidique marquée à son extrémité par du $g^{32}P$ ATP par action de la T4 polynucléotide kinase. Après préhybridation de la membrane dans une solution de 20 ml contenant 6 ml de SSC 20X, 1 ml de solution de Denhardt, 1 ml de SDS 10% et 5 mg d'ADN hétérologue de sperme de saumon, les hybridations sont conduites durant une nuit à la température définie par la sonde. Les membranes sont lavées deux fois dans du SSC 2X pendant 5 minutes à température ambiante, puis une fois dans du SSC 2X SDS 0,1% et une seconde fois dans du SSC 1X, SDS 0,1% pendant 30 minutes à la température d'hybridation. Les signaux d'hybridation sont quantifiés à l'aide du logiciel Molecular Analyst (Biorad, Ivry sur Seine, France) et les quantités d'ADN sont estimées par interpolation des courbes étalons obtenues à partir des ADN génomiques.

2. RESULTATS ET DISCUSSION

2.1 Extraction et lyse de la fraction bactérienne du sol.

La séparation des cellules microbiennes des particules du sol, préalablement à l'extraction de l'ADN, est une alternative présentant de nombreux avantages par rapport aux méthodes d'extraction directe de l'ADN dans le sol. En effet, l'extraction de la fraction microbienne limite la contamination de l'extrait d'ADN par de l'ADN extracellulaire présent librement dans le sol ou par de l'ADN d'origine eucaryote. Mais surtout, l'ADN extrait de la fraction microbienne du sol présente des fragments de plus longue taille et une meilleure intégrité que l'ADN extrait par lyse directe JACOBSON et RASMUSSEN (1992). De plus, la séparation des

particules de sol permet d'éviter une contamination de l'extrait d'ADN par des composés humiques et phénoliques, composés pouvant, par la suite, nuire gravement aux efficacités de clonage.

Une des étapes déterminantes pour l'extraction des cellules du sol est la dispersion de l'échantillon de sol afin de dissocier les cellules adhérent à la surface ou à l'intérieur des agrégats de particules de sol. Trois cycles de broyage successifs d'une minute chacun permettent d'obtenir une meilleure efficacité d'extraction des cellules ainsi qu'une plus grande quantité d'ADN récupéré, par rapport à un unique cycle de broyage d'une minute 30.

Le tableau 5 rapporte les efficacités d'extraction obtenues après centrifugation sur gradient de Nycodenz, sur la microflore totale viable (dénombrée par microscopie après coloration à l'acridine orange), sur la microflore totale cultivable (dénombrée sur milieu solide Trypticase-Soja 10%), et sur la microflore d'actinomycètes cultivables sur milieu HV agar (après incubation à 40°C dans une solution d'extrait de levure 6% -SDS 0,05% afin de provoquer la germination des spores). D'autre part, l'ADN extrait a été quantifié soit après une lyse des cellules en milieu liquide (sans purification sur gradient de chlorure de césium) soit après une lyse des cellules incluses dans un bloc d'agarose (après digestion de l'agarose par une α -agarase).

Les résultats montrent que plus de 14% de la microflore tellurique totale est récupéré par cette méthode (soit 2×10^8 cellules par gramme de sol), et que la microflore totale cultivable ne représente qu'à peine 2% de la population microbienne totale.

D'autre part, la quantité d'ADN extrait des cellules est de 330 ng par gramme de sol sec. En estimant le contenu d'ADN par cellule microbienne du sol entre 1.6 et 2.4 fg, et compte tenu de la quantité de cellules extraites (2×10^8 cellules par gramme de sol), on peut estimer que la quasi-totalité des cellules ont été lysées et qu'ainsi la lyse n'apporte pas d'important biais à cette approche.

Les électrophorèses en champs pulsés ont montré que l'ADN du sol extrait après gradient de Nycodenz et de CsCl pouvait atteindre une taille de 150 kb et que la lyse en bloc d'agarose permettait d'extraire des fragments supérieurs à 600kb.

Ces résultats confirment l'intérêt de cette approche indépendante de la culture pour la construction de banques d'ADN de l'environnement, en se présentant comme une alternative aux méthodes directes d'extraction d'ADN.

5

2.2 Caractérisation moléculaire de l'ADN extrait du sol.

Le but de la caractérisation moléculaire de l'ADN extrait du sol est d'obtenir des profils représentant les proportions des différents
10 taxons bactériens présents dans l'extrait d'ADN. Il s'agissait également de connaître les biais d'extraction induits par la séparation préalable de la réaction cellulaire du sol, en comparaison avec une méthode d'extraction directe faute de visualisation directe de la diversité microbienne présente dans les sols. En effet, peu d'informations ont été
15 rassemblées sur l'extraction des cellules sur gradient de Nycodenz en fonction de leur structure morphologique (diamètre des cellules, formes filamenteuses ou sporulées).

Les méthodes jusqu'ici en place étaient basées sur des:

- hybridations quantitatives utilisant des sondes
20 oligonucléotidiques spécifiques à différents groupes bactériens, appliqués directement d'ADN extrait de l'environnement. Malheureusement, cette approche n'est pas très sensible et ne permet pas de détecter des genres ou des groupes taxonomiques présents en faible abondance AMANN (1995).

25 - PCR quantitatives telles que la MPN-PCR (Most Probable Number) SYKES et al. (1992) ou la PCR quantitative par compétition DIVIACCO et al. (1993). Les inconvénients respectifs de chacune de ces approches sont (i) la lourdeur d'utilisation du fait de la multiplication des dilutions et des répétitions qui rend la technique inappropriée pour un
30 grand nombre d'échantillons ou de couples d'amorces, et (ii) la nécessité de construire un compétiteur spécifique à l'ADN cible et n'induisant pas de biais dans la compétition.

La méthode mise en place selon la présente invention consiste à amplifier universellement un fragment de 700 pb à l'intérieur de la
35 séquence d'ADNr 16S, à hybrider cet amplifiat avec une sonde

oligonucléotidique de spécificité variable (au niveau du règne, de l'ordre, de la sous classe ou du genre) et à comparer l'intensité d'hybridation de l'échantillon par rapport à une gamme étalon externe. L'amplification préalable à l'hybridation permet de quantifier des genres ou des espèces de micro-organismes peu abondants. De plus, l'amplification par des amorces universelles permet, lors de l'hybridation, d'utiliser une large série de sondes oligonucléotidiques. Elle permet de comparer entre eux différents modes de lyse (extraction directe ou indirecte) sur des groupes taxonomiques bien définis.

10 Les résultats sont rassemblés dans le tableau 6.

Ils montrent des profils similaires entre les deux méthodes d'extraction (directe et indirecte). Ainsi, il apparaît que l'extraction préalable de la fraction microbienne tellurique n'introduit pas de réels biais parmi les taxons testés. La seule différence significative entre les deux approches d'extraction semblerait être la plus grande abondance de séquences d'ADNr appartenant aux γ protéobactéries dans l'extrait par la méthode d'extraction indirecte.

De plus, un effet significatif de l'incubation de l'échantillon de sol dans une solution d'extrait de levure est observé sur les populations sporulées du sol (Gram⁺, bas pourcentage de GC et Actinomycètes). Cette étape provoque la germination des spores, et permet d'une part certainement une meilleure récupération de ce type de cellules et d'autre part une plus grande efficacité de la lyse sur des cellules en germination.

Cette approche permet une analyse semi-quantitative, ciblée sur les principaux taxons définis à partir de micro-organismes cultivés et habituellement retrouvés dans les sols. Seuls des outils moléculaires permettent d'estimer l'importance des différents taxons, les méthodes de mise en culture étant trop restrictives et dépendantes de la spécificité du milieu utilisé.

30 Les résultats montrent qu'une grande part de la population microbienne n'est pas représentée dans les groupes phylogénétiques décrits, mettant ainsi en évidence l'existence de nouveaux groupes composés de micro-organismes non cultivés jusqu'à présent, ou non cultivables.

Ainsi, de nouvelles sondes peuvent être définies à partir de séquences déterminées à partir d'ADN extrait du sol (nouveaux phylums composés de micro-organismes non cultivés, LUDWIG et al. (1997) afin d'obtenir une image plus exacte de la composition de l'extrait d'ADN.

5

Exemple 4 : - CONSTRUCTION DU COSMIDE POS 700I

Caractéristiques de POS 700I:

- Réplicatif chez *E. coli*
- 10 Intégratif chez *Streptomyces*
- Sélectionnable chez *E. coli* AmpR, HygroR et *Streptomyces* HygroR

Les propriétés du cosmide permettent d'insérer de grands fragments d'ADN entre 30 et 40kb.

15 Il comprend

- 1 - Le promoteur inductible *tipA* de *Streptomyces lividans*
- 2 - Le système d'intégration spécifique de l'élément pSAM2
- 3 - Le gène de résistance à l'hygromycine
- 4- le cosmide pWED1, dérivé de pWED15

20

1) - Le promoteur inductible du gène *tip A* de *S. lividans*

Le gène *tipA* code une protéine de 19 KD dont la transcription est induite par l'antibiotique thiostrepton ou nosiheptide. Le promoteur de
25 *tipA* est bien régulé: induction en phase exponentielle et en phase stationnaire (200X) Murakami T, Holt TG, Thompson CJ. J. Bacteriol 1989 ;171 :1459-66

2) - Le gène de résistance à l'hygromycine

30

- Hygromycine: antibiotique produit par *S. hygroscopicus*
- Le gène de résistance code une phosphotransférase (*hph*)
- Le gène utilisé provient d'une cassette construite par Blondelet et al dans laquelle le gène *hyg* est sous contrôle de son propre promoteur

et du promoteur plac inducible par l'IPTG Blondelet-Rouault et al ; .
Gene 1997 ;190 :315-7

3) - Le système d'intégration site-spécifique

5

L'élément pSAM2 s'intègre dans le chromosome par un mécanisme d'intégration site-spécifique. La recombinaison a lieu entre deux séquences identiques de 58 pb présentes sur le plasmide (*attP*) et sur le chromosome (*attB*).

10 Le gène *int*, situé à proximité du site *attP*, est impliqué dans l'intégration site-spécifique de pSAM2, et son produit présente des similitudes avec les intégrases des bactériophages tempérés d'entérobactéries. Il a été démontré qu'un fragment de pSAM2 ne
15 contenant que le site d'attachement *attP* ainsi que le gène *int* était capable de s'intégrer de la même manière que l'élément entier. Voir brevet français n°88 06638 du 18/05/1988, ainsi que Raynal A et al. Mol Microbiol 1998 28 :333-42).

4) – Construction du cosmide pOS700I

20

Etape 1/ Le promoteur TipA a été isolé du plasmide pPM927 (Smokvina et al. Gene 1990; 94:53-9) sur un fragment HindIII-BamHI de 700 paires de bases et cloné dans le vecteur pUC18 (Yannish-Perron et al., 1985) digéré par HindIII/BamHI

25

Etape 2/ Ce fragment HindIII-BamHI a ultérieurement été transféré de pUC18 à pUC19 (Yannish-Perron et al., 1985).

Etape 3/ Un insert BamHI-BamHI de 1500 paires de bases portant le
30 gène *int* et le site *attP* de pSAM2 a été isolé du plasmide pOSint1, représenté à la Figure 8, (Raynal A et al. Mol Microbiol 1998 28 :333-42) et cloné au site BamHI du vecteur précédent (pUC19/TipA), dans l'orientation permettant de mettre le gène *int* sous contrôle du promoteur TipA.

35

Etape 4/ Le site BamHI situé en 5' du gène int a été supprimé par digestion partielle BamHI puis traitement par l'enzyme Klenow. Un fragment HindIII-BamHI portant TipA-int-attP a ainsi été isolé de pUC19 et transféré dans pBR322 HindIII/BamHI.

5

Etape 5/ La cassette Hygromycine isolée de pHP45 Ω hyg (Blondelet-Rouault et al., 1997) sur un fragment HindIII-HindIII a été clonée au site HindIII situé en amont du promoteur TipA.

10 **Etape 6/** Le site HindIII situé entre la cassette Ω Hyg et le promoteur TipA a été supprimé par traitement Klenow après digestion partielle HindIII.

15 **Etape 7/** Le plasmide obtenu à l'issue de l'étape précédente permet d'isoler un fragment unique HindIII-BamHI, portant tous les éléments Ω Hyg/TipA/int attP, qui a été cloné après traitement Klenow au site EcoRV du cosmide pWED1. Le cosmide pWED1, représenté à la Figure 9, dérive du cosmide pWE15, représenté à la Figure 10 (Wahl GM, et al. Proc Natl Acad Sci U S A 1987 84:2160-4) par délétion d'un fragment
20 HpaI-HpaI portant le gène Neomycine et l'origine SV40.

Une carte du vecteur pOS 700I est représentée à la Figure 11.

25 **Exemple 5 : Construction de plusieurs cosmides conjugatifs et intégratifs chez *Streptomyces*, les vecteur pOSV 303, pOSV306 et pOSV307**

5.1 Construction du vecteur pOSV303.

30 Etant donné que l'empaquetage sélectionne les clones ayant une taille supérieure à 30kb, seuls 10 à 15% des clones ne contiennent pas d'insert, il n'est donc pas vraiment nécessaire d'avoir un système de sélection des recombinants, ce qui permet de construire un vecteur plus petit.

35

Construction:**Etape 1 : le vecteur pOSV001**

Clonage d'un fragment PstI-PstI de 800 paires de bases portant l'origine de transfert OriT du réplicon RK2 (Guiney et al., 1983), dans le plasmide pUC19 ouvert par PstI. Cette étape de clonage permet d'obtenir un vecteur transférable de *E. coli* à *Streptomyces* par conjugaison.

La carte du vecteur pOSV 001 est représentée à la Figure 17.

10 Etape 2 : le vecteur pOSV002

Insertion du marqueur Hygromycine (cassette Ω hyg), et sélectionnable chez *Streptomyces*, de sorte que le gène conférant la résistance à l'hygromycine soit transféré en dernier ce qui permet de s'assurer du transfert complet du BAC avec l'insert d'ADN du sol.

15 Clonage de la cassette Hygromycine isolée de pHP45 Ω hyg sur un fragment HindIII-HindIII portant le gène de résistance à l'Hygromycine.. Ce fragment est cloné au site PstI (position 201) du vecteur pOSV001. Ce site PstI a été choisi, compte tenu du sens du transfert, pour que le marqueur Hygro soit le dernier transféré lors de la conjugaison. Les extrémités PstI et HindIII sont rendues compatibles après traitement par le fragment Klenow de l'ADN polymérase permettant de générer des "bouts francs". L'orientation du fragment Ω hyg est déterminée en fin de construction.

La carte du vecteur pOSV002 est représentée à la Figure 18.

25

Etape 3 : le vecteur pOSV010

Le fragment XbaI-HindIII isolé du plasmide pOSV002 et contenant le marqueur de résistance à l'hygromycine et l'origine de transfert est cloné dans le plasmide pOSint1 digéré par XbaI et HindIII. L'orientation des sites est telle que le marqueur hygromycine sera toujours transféré en dernier.

Le plasmide pOSint1, représenté à la Figure 8, a été décrit dans l'article de Raynal et al. (Raynal A et al. Mol Microbiol 1998 28 :333-42).

35 Cette construction permet l'expression de l'intégrase chez *E. coli* et chez *Streptomyces*.

Etape 4 : insertion du site " cos "

Le principe est d'insérer un site " cos " dans le plasmide pOSV010 permettant l'empaquetage dans le plasmide pOSV010, représenté à la Figure 12.

L'obtention du fragment " cos " est représentée à la Figure 13.

Ce fragment est obtenu par PCR. A partir d'un fragment portant les extrémités cohésives (cos) de λ (bactériophage lambda ou cosmide pHc79), une amplification par PCR est réalisée à l'aide des oligonucléotides correspondant aux séquences -50/+130 par rapport au site cos. Ces oligonucléotides contiennent en outre les sites de clonage NsiI, compatible PstI, XhoI, compatible Sall, EcoRV, " bout franc ".

L'addition des sites rares Swal et PacI permet d'isoler et/ou de cartographier l'insert cloné.

Le fragment PCR est borné par un site PstI à l'extrémité 5' et par un site HincII à l'extrémité 3', permettant le clonage dans le vecteur pOSV010 (Figure 12) préalablement digéré par les enzymes NsiI et EcoRV, provoquant la délétion du répresseur lacIq.

La carte du vecteur pOSV303 est représentée sur la Figure 14. Le vecteur pOSV303, contient des sites de clonage tels que le site NsiI, compatible PstI, le site XhoI, compatible Sall ou encore le site EcoRV pour l'obtention de " bouts francs ".

5.2 Construction du vecteur pOSV306**Etape 1: Construction du vecteur pOSV308.**

Le vecteur pOSV308 a été construit selon le procédé illustré à la figure 27. Un fragment de 643 pb contenant la région cos a été amplifié à l'aide du couple d'amorces de séquences SEQ ID N°107 et SEQ ID N°108 à partir du vecteur cosmide pHc79 décrit par HOHM B and COLLINS (1980).

Ce fragment nucléotidique amplifié a été cloné directement dans le vecteur pGEMT-easy commercialisé par la Société PROMEGA, comme illustré à la figure 27 afin de produire le vecteur pOSV308.

5 **Etape 2: Construction du vecteur pOSV306.**

Le vecteur pOSV010 a été construit comme décrit à l'étape 3 de construction du vecteur pOSV303, comme décrit au paragraphe 5.1 du présent exemple.

10 Le vecteur pOSV10 a été digéré par les enzymes EcoRV et NsiI afin d'exciser un fragment de 7874 pb qui a été ultérieurement purifié, comme cela est illustré à la figure 28.

Puis, le vecteur pOSV308 obtenu à l'étape 1) ci-dessus a été soumis à une digestion par les enzymes EcoRV et PstI afin d'exciser un
15 fragment de 617 pb, qui a été ultérieurement purifié.

Puis, le fragment cos de 617 pb obtenu à partir du vecteur pOSV308 a été intégré par ligation dans le vecteur pOSV10, afin d'obtenir le vecteur pOSV306, comme cela est illustré à la figure 28.

20 **5.3 Construction du vecteur pOSV307.**

Le cosmide pOSV307 contient toujours le gène LacIq, afin d'améliorer la stabilité du cosmide dans *Streptomyces*, par exemple dans la souche S17-1 de *Streptomyces*.

25 Afin de construire le vecteur pOSV307, on a soumis le vecteur pOSV010 à une digestion par l'enzyme PvuII, pour obtenir un fragment de 8761 pb qui a été purifié, puis déphosphorylé.

Ensuite, le vecteur pOSV308, tel qu'obtenu comme décrit à l'étape 1) du paragraphe 5.2 ci-dessus, a été digéré par l'enzyme EcoRI
30 afin d'obtenir un fragment de 663 pb, qui a été ensuite purifié et traité par l'enzyme de Klenow.

Le fragment nucléotidique ainsi traité a été intégré dans le vecteur pOSV010 après ligation afin d'obtenir le vecteur pOSV307, comme illustré à la figure 29.

Exemple 6 : - Construction du cosmide répliatif navette *E. coli*-*Streptomyces* pOS700R.

Les fragments du plasmide pEI16 (Voff et al., 1996) représenté
5 à la Figure 15 ont été isolés et traités par Klenow. Ces fragments
contiennent les séquences nécessaires à la répliation et à la stabilité
provenant du plasmide SCP2.

Ces deux fragment sont insérés séparément dans le site
EcoRV du cosmide pWED1 conduisant à 2 clones différents.

10 La cassette Hygromycine isolée de pHP45Ωhyg sur un
fragment HindIII-HindIII a été clonée au site HindIII des cosmides
pWED1 contenant l'insert ScP2 sous forme de fragments PstI-EcoRI ou
XbaI. Elle confère une résistance à l'Hygromycine sélectionnable à la
fois chez *E. coli* et chez *Streptomyces*.

15 Transformation de *S. lividans* et détermination de l'efficacité de
transformation.

Il est apparu que le cosmide contenant l'insert XbaI était moins
stable que celui contenant le fragment PstI EcoRI. C'est donc ce dernier
qui a été retenu sous le nom de pOS700R.

20 La carte du vecteur pOS 700R est représentée sur la Figure 16.

**Exemple 7: Efficacité de transformation des vecteurs intégratifs
(pOS700I) et répliatifs**

25 **Possibilités**

Rendre la souche de *S. lividans* résistante au thiostrepton par
intégration du plasmide pTO1 portant le marqueur de résistance au
thiostrepton

30 Préparation de protoplastes à partir de *S. lividans* cultivée en
présence de thiostrepton

Avec le vecteur pOS700I, l'efficacité de transformation est
d'environ 3000 transformants par µg d'ADN.

Avec le vecteur pOS700R, l'efficacité de transformation est
d'environ 30 000 transformants par µg d'ADN.

Exemple 8 : Construction d'un vecteur BAC intégratif chez *Streptomyces* et conjugatif

Caractéristiques:

5

Réplicatif chez *E. coli*

Transférable par conjugaison de *E. coli* aux *Streptomyces*

Intégratif chez *Streptomyces*

Sélectionnable chez *E. coli* et *Streptomyces*

- 10 Capable d'insérer de grands fragments d'ADN ; il faut souligner qu'il est nécessaire de disposer d'ADN du sol dont la taille est comprise entre 100 et 300kb et non contaminé par des petits fragments. En effet les petits fragments sont très préférentiellement intégrés.

- 15 Doté d'un crible permettant de sélectionner les plasmides portant un insert. Ce crible permet en éliminant les vecteurs refermés sur eux même et non digérés de travailler avec un rapport plus élevé entre vecteur et DNA à insérer ce qui permet d'avoir une meilleure efficacité de clonage pour constituer des banques.

20 **Construction:**

Etape 1 : le vecteur pOSV001

- 25 Clonage d'un fragment PstI-PstI de 800 paires de bases portant l'origine de transfert OriT du réplicon RK2 (Guiney et al., 1983), dans le plasmide pUC19 ouvert par PstI. Cette étape de clonage permet d'obtenir un vecteur transférable de *E. coli* à *Streptomyces* par conjugaison.

La carte du vecteur pOSV 001 est représentée à la Figure 17.

30

Etape 2 : le vecteur pOSV002

Insertion du marqueur Hygromycine (cassette Ω hyg), et sélectionnable chez *Streptomyces*, de sorte que le gène conférant la

résistance à l'hygromycine soit transféré en dernier ce qui permet de s'assurer du transfert complet du BAC avec l'insert d'ADN du sol.

Clonage de la cassette Hygromycine isolée de pHP45Ωhyg sur un fragment HindIII-HindIII portant le gène de résistance à l'Hygromycine..

5 Ce fragment est cloné au site PstI (position 201) du vecteur pOSV001. Ce site PstI a été choisi, compte tenu du sens du transfert, pour que le marqueur Hygro soit le dernier transféré lors de la conjugaison. Les extrémités PstI et HindIII sont rendues compatibles après traitement par le fragment Klenow de l'ADN polymérase permettant de générer des
10 "bouts francs". L'orientation du fragment Ωhyg est déterminée en fin de construction.

La carte du vecteur pOSV002 est représentée à la Figure 18.

Etape 3 : le vecteur pOSV010

15

Le fragment XbaI-HindIII isolé du plasmide pOSV002 et contenant le marqueur de résistance à l'hygromycine et l'origine de transfert est cloné dans le plasmide pOSint1 digéré par XbaI et HindIII. L'orientation des sites est telle que le marqueur hygromycine sera
20 toujours transféré en dernier.

Le plasmide pOSint1, représenté à la Figure 8, a été décrit dans l'article de Raynal et al.(Raynal A et al. Mol Microbiol 1998 **28** :333-42).

25

Cette construction permet l'expression de l'intégrase chez *E. coli* et chez *Streptomyces*.

Etape 4 : le vecteur pOSV014

30

Addition d'une "cassette" permettant à terme de sélectionner dans la construction finale les plasmides ayant insérés de l'ADN étranger.

Cette "cassette" porte le gène codant pour le répresseur CI du phage λ et le gène conférant la résistance à la tétracycline. Ce gène porte dans
35 sa région 5' non codante la séquence cible du répresseur. L'insertion

d'ADN dans le site HindIII situé dans la séquence codante de CI conduit à la non production du répresseur et donc à l'expression de la résistance à la tétracycline.

Elle est portée par le plasmide pUN99 décrit dans l'article : Nilsson et al .

5 (Nucleic Acids Res 1983, 11:8019-30)

Un fragment PvuII-HindIII isolé de pOSV010 et contenant les séquences Int, attP, Hygro et oriT est cloné au site MscI de pUN99 .

La carte du vecteur pOSV014 est représentée sur la Figure 19.

10 **Etape 5 : le vecteur pOSV 403, vecteur BAC intégratif et conjugatif**

Cette dernière étape de clonage dans pBAC11 (représenté à la Figure. 20) permet de conférer au plasmide final des caractéristiques de BAC (Bacterial Artificial Chromosome), en particulier l'aptitude à
15 accepter des inserts d'ADN de très grande taille.

Le fragment PstI-PstI du vecteur pOSV014 portant l'ensemble des éléments et fonctions décrits précédemment est cloné dans le plasmide pBAC11 (pBeloBAC11) digéré par NotI. Les extrémités sont rendues compatibles par traitement avec l'enzyme de Klenow.

20 La carte du vecteur pOSV403 est représentée sur la Figure 21. Le schéma de la Figure 21 indique l'orientation retenue.

Etape 6 :

Le vecteur pOSV403 contient les sites HindIII et NsiI. Le site
25 NsiI est assez rare chez *Streptomyces* et présente l'avantage d'être compatible avec PstI. En revanche, le site PstI est fréquent chez *Streptomyces* et peut être utilisé pour effectuer des digestions partielles.

Les clones recombinants portant un insert cloné dans le répresseur CI, et donc inactivant ce répresseur deviennent résistants à
30 la tétracycline. Etant donné que les BACs ne sont présents qu'à raison d'une copie par cellule, il faut sélectionner les clones recombinants avec une dose plus faible de tétracycline que la dose habituelle de 20 µg/ml, par exemple avec une dose de 5 µg/ml. Dans ces conditions il n'y a aucun bruit de fond.

Il est aussi possible d'utiliser un système développé et commercialisé par la société InVitrogen, dans lequel l'insertion d'ADN dans le vecteur inactive un inhibiteur de la gyrase dont l'expression est toxique pour *E. coli*. Le fragment est préférentiellement isolé à partir du vecteur pZErO-2 (<http://www.invitrogen.com/>).

Exemple 9 : Construction d'une banque de *S. alboniger* dans les 2 cosmides intégratif (pOS700I) et réplicatif (pOS700R)

1) - Construction de la Banque

Pour évaluer l'efficacité du système de clonage, la voie de biosynthèse de la puromycine de *Streptomyces alboniger*, a été clonée dans les deux cosmides navettes pOS700I et pOS700R. Les gènes de la voie de biosynthèse de la puromycine sont portés par un fragment d'ADN BamHI d'environ 15 Kb.

L'ADN génomique de *Streptomyces alboniger* a été isolé. 90% de cet ADN possède un poids moléculaire compris entre 20 et 150 Kb, déterminé par électrophorèse en champ pulsé.

Les deux cosmides ont été digérés par l'enzyme *Bam*HI (site unique de clonage).

Les conditions de digestion partielle *Bam*HI de l'ADN génomique ont été déterminées (50 µg d'ADN et 12 unités d'enzyme, 5 minutes de digestion). Après vérification de la taille par électrophorèse en gel d'agarose, l'ADN partiellement digéré a été introduit dans les vecteurs. Dans la ligation, 15 µg d'ADN génomique + 2 µg du vecteur intégratif ou 5 µg du vecteur réplicatif ont été utilisés.

Chaque mélange de ligation a été utilisé pour l'encapsidation in vitro de l'ADN dans les têtes de bactériophage lambda. Les mélanges d'encapsidation (0,5ml) ont été titrés (Vecteur intégratif pOS700I = $7,5 \times 10^5$ cosmides/ml, Vecteur réplicatif = 5×10^4 cosmides/ml).

Les cosmides ont été utilisés pour transfecter *E. coli* et générer ainsi deux banques d'environ 25000 clones résistant à l'ampicilline. L'ADN de l'ensemble de ces clones a été isolé et quantifié.

Pour tester les banques, plusieurs clones ont été choisis, l'ADN purifié et a été digéré par BamHI, afin de vérifier la présence et la taille des inserts. Les clones testés contiennent entre 20 et 35 Kb d'insert de *S. alboniger*.

5

2) - Identification des clones contenant la voie de biosynthèse de la puromycine

Les clones susceptibles de contenir la voie complète de biosynthèse de la puromycine ont été identifiés par hybridation avec une sonde correspondant au gène de résistance à la puromycine, le gène *pac* de 1,1 kb. (Lacalle et al. Gene 1989;79, 375-80)

10

Banque faite dans le Vecteur Intégratif pOS 700I:

15

Parmi 2000 clones analysés, 9 clones ont hybridé avec la sonde et ils contiennent des inserts d'environ 40 kb.

Banque faite dans le Vecteur replicatif pOS 700R:

20

Parmi 2000 clones analysés, 12 clones ont hybridé avec la sonde; ils contiennent des inserts d'environ 40 kb.

25

En utilisant les données publiées par Tercero et al. (J Biol Chem. 1996; 271, 1579-90), les clones contenant la totalité de la voie de biosynthèse ont été identifiés, après hybridation avec des sondes appropriées. Certains cosmides intégratifs et replicatifs présentent après digestion ClaI-EcoRV un fragment de 12360 paires de bases, ce qui laisse supposer un insert contenant la totalité de la voie de biosynthèse de la puromycine.

30

4) - Vérification de la production de puromycine par les clones résistants (Rhône-Poulenc).

a) Matériels et Méthodes**Souches et conditions de culture :**

5

Trois clones résistants ont été sélectionnés pour vérifier la production de puromycine. Ils correspondent aux recombinants de *S. lividans* contenant un insert dans le vecteur intégratif pOS700I (G 20) ou un insert dans le vecteur réplcatif (G21 et G22).

10

Des souches de référence ont été utilisées pour s'assurer que les milieux de culture utilisés permettaient cette production. Il s'agit de la souche sauvage *S. alboniger* ATCC 12461, productrice de puromycine et de la souche recombinante *S. lividans* contenant le cluster complet de la puromycine cloné dans le plasmide pRCP11 (Lacalle et al, 1992, the EMBO journal, 11, 785-792) (G23).

15

Les souches sontensemencés dans un milieu de culture dont la composition est la suivante :

20	Peptone bactériologique Organotechnie	5 g/l de milieu final
	Extrait de levure Springer	5
	Extrait de viande Liebig	5
	Glucose Prolabo	15
	CaCO ₃ (1) Prolabo	3
25	NaCl Prolabo	5
	Agar (2) Difco	1

(1) Les 3g de carbonate sont mélangés à 200ml d'eau distillée puis stérilisés à part. L'addition se faisant après stérilisation.

30 (2) L'agar est préalablement fondu dans 100ml d'eau distillée avant d'être ajouté aux autres ingrédients du milieu

pH ajusté à 7,2 avant stérilisation
stérilisation 25 minutes à 121°C

35

50 µg/l d'hygromycine et 5 µg/l de thiostrepton sont ajoutés au milieu après stérilisation de façon à maintenir une pression de sélection des clones contenant un insert grâce au gène marqueur présent sur le vecteur (le gène de résistance au thiostrepton étant porté par le plasmide pRCP11).

50 ml de milieu de culture liquide, répartis en erlenmeyers de 250 ml, sont ensemencés avec 2 ml de suspension aqueuse de spores et de mycelium de chacune des souches. Les cultures sont incubées pendant 4 jours à 28°C avec une agitation de 220 trs/mn. 50 ml de milieux de production, répartis en erlenmeyers de 250 ml, sont ensuite ensemencés avec 2 ml de ces pré-cultures. Le milieu de production utilisé est un milieu industriel optimisé pour la production de pristinamycine (milieu RPR 201). Les cultures sont incubées à 28°C, avec une agitation de 220trs/mn. Après différents temps d'incubation, un erlenmeyer de chaque culture est amené à pH 11 puis extrait par 2 fois 1 volume de dichlorométhane. La phase organique est concentrée à sec sous pression réduite, puis l'extrait est repris par 10 µl de méthanol. 100 µl de la solution méthanolique sont analysés en CLHP munie d'un détecteur à barrette de diodes dans un système gradient eau-acétonitrile 0,05% TFA V/V sur colonne C18 pour la détection de la puromycine.

b) Résultats

Les analyses HPLC comparatives à partir des cultures des différentes souches montrent la production de puromycine dans la culture de la souche sauvage à partir de 24 h d'incubation. Une production, bien que plus faible, est aussi nettement détectée à partir de 48 h dans la culture du clone G20 contenant le cosmide pOS700I (figure 23). La puromycine a également été détecté à l'état de trace dans le clone G23 contenant l'opéron complet codant pour le composé dans le plasmide pRCP11. Néanmoins, aucune production n'a été observée dans les cultures des clones G21 et G22 contenant le cosmide pOS700R. Les résultats sont reportés sur la Figure 23.

c) Conclusions

Les résultats obtenus permettent de démontrer l'efficacité du système de clonage développé dans le cosmide pOS700I pour exprimer chez un hôte hétérologue tel que *S. lividans* une voie de biosynthèse complète sous le contrôle de séquences régulatrices qui lui sont propres. D'autre part, ces données valident également le criblage des banques obtenues sur la base de la résistance des clones à la puromycine puisqu'il a conduit à identifier parmi un petit nombre de clones, un recombinant capable d'exprimer la voie de biosynthèse associée au gène de résistance. L'absence de production de puromycine chez les autres clones peut probablement s'expliquer par le clonage d'une partie seulement de l'opéron contenant le gène de résistance mais dépourvue de certaines séquences de régulation, transduction ou transcription nécessaires à la synthèse du composé.

EXEMPLE 10 : - CLONAGE D'ADN DU SOL DANS DES VECTEURS

1) – Préparation de l'ADN du sol à cloner

Les différents fragments d'ADN doivent être purifiés selon leur destination :

Cosmides

La taille des molécules doit être comprise entre 30 et 40kb. Or, l'ADN extrait du sol est hétérogène en taille et comprend des molécules atteignant 200 ou 300kb. Afin d'homogénéiser les tailles, l'ADN est cassé mécaniquement par passage de la solution à travers une aiguille de 0,4mm de diamètre. Les fragments d'une taille voisine de 30kb ne sont pas affectés par ces passages répétés à travers une aiguille et il n'est donc pas nécessaire de faire une séparation par la taille surtout que l'emballage dans les particules élimine automatiquement les inserts courts.

BACs

Préparation de l'ADN

5 L'ADN du sol est séparé par électrophorèse en champ pulsé (type CHEF) dans des conditions telles que les fragments compris entre 100 et 300kb sont concentrés dans une bande d'environ 5mm. Ceci est obtenu en réalisant la migration dans un gel à 0,7% d'agarose normal ou 1% d'agarose à bas point de fusion avec un temps de pulsation de 100 secondes pendant 20 heures et à une température
10 de 10°C.

Récupération de l'ADN

15 Deux méthodes sont utilisées, leur choix dépend de la taille des molécules que l'on veut isoler, soit jusqu'à 150kb soit au dessus.

- Jusqu'à 150kb

20 La porosité d'un gel à 0,7% d'agarose permet la sortie de l'ADN par électroélution à condition d'absence totale de bromure d'ethidium. Cet ADN est ensuite manipulé avec des instruments de pipetage à orifice agrandi et hydrophobe pour éviter la fragmentation mécanique des molécules.

- Entre 100 et 300kb

25 La bande contenant les fragments d'une taille entre 100 et 300kb est découpée. Pour la migration un gel d'agarose à 1% et à bas point de fusion est utilisé. Cette propriété permet de fondre le gel à une température supportable pour l'ADN de 65°C et de le digérer ensuite
30 par l'agarase (Agarase commercialisée par la société Boehringer) à une température de 45°C suivant les prescriptions du fournisseur.

2) - Utilisation des cosmides intégratifs pOS700I et replicatifs pOS700R

5 Construction par queues polyA polyT

Principe

Un vecteur cosmide, ouvert à un site de clonage quelconque, est modifié aux extrémités 3' en ajoutant un polynucléotide monotone. D'autre part,
10 l'ADN à cloner est modifié aux extrémités 3' en ajoutant un polynucléotide monotone pouvant s'apparier au précédent.

L'association vecteur-fragment à cloner se fait par ces polynucléotides et la séquence cos du vecteur permet l'empaquetage *in vitro* de l'ADN dans
15 des capsides de phage Lamda.

Préparation du vecteur

Le vecteur utilisé est un vecteur autorépliquatif chez *E. coli* et intégratif
20 chez *Streptomyces*.

Pour *E. coli*, la sélection se fait sur la résistance à l'ampicilline et pour *Streptomyces*, elle se fait sur la résistance à l'hygromycine .

Le cosmide est ouvert à l'un des 2 sites possibles (BamHI ou HindIII) et
25 les extrémités 3' sont rallongées par du polyA avec de la terminale transférase dans les conditions où le fournisseur de l'enzyme prévoit l'addition de 50 à 100 nucléotides.

Préparation de l'ADN à insérer.

30 Les extrémités 3' de l'ADN sont rallongées par du polyT avec de la terminale transférase dans les conditions fournissant un allongement comparable à celui du vecteur. Dans les conditions expérimentales décrites par le fabricant les queues polyA polyT sont longues de 30 à 70
35 bases

Assemblage des molécules et encapsidation in vitro.

Pour l'assemblage des molécules, on mélange une molécule de vecteur
5 pour une molécule d'ADN inséré. La concentration de l'ADN en masse
est de $500 \mu\text{g}.\text{ml}^{-1}$.

Le mélange est encapsidé et l'efficacité de transfection dépend de la
souche utilisée comme réceptrice et de l'ADN inséré : nulle avec l'ADN
test et la souche DH5 α , l'efficacité est comparable pour les souches
10 SURE et DH10B ; à l'extraction le rendement en ADN est cependant
plus élevé avec la souche DH10B.

Construction par déphosphorylation

15 L'ADN du sol est mis en bouts francs par élimination des séquences 3'
sortantes et remplissage des séquences 5' sortantes. Cette opération est
faite avec : enzyme de Klenow, T4 polymérase, les 4 nucléotides
triphosphates. Le vecteur cosmétique est digéré par BamHI, puis traité
par l'enzyme de Klenow pour le rendre bout franc puis déphosphorylé
20 pour éviter qu'il ne se referme sur lui même. Après ligation, le mélange
est encapsidé et transfecté comme précédemment décrit.

3) - Utilisation des pBAC

Principe.

25

Le plasmide pBAC conjugatif et intégratif possède les sites HindIII et NsiI
comme sites de clonage. L'insertion d'une séquence d'ADN à ces sites
inactive le répresseur CI du phage Lambda qui contrôle l'expression du
gène de la résistance à la tétracycline. L'inactivation du répresseur rend
30 donc la cellule résistante à cet antibiotique ($5\mu\text{g}.\text{ml}^{-1}$). Le clonage à ces
sites est facilité par la modification du vecteur et la préparation de l'ADN
à cloner.

Préparation du vecteur. Exemple HindIII

Pour que le vecteur ne se referme pas sur lui-même, le site Hind III est
5 modifié : la première base (A) est remise en place pour former une
séquence 5' sortante, qui ne peut pas s'apparier avec ses semblables.
L'opération est effectuée par l'enzyme de Klenow en présence de dATP.

Le succès de l'opération est vérifié en effectuant une ligation du vecteur
10 sur lui-même avant et après traitement à l'enzyme de Klenow. A quantité
d'ADN testé identique, on obtient 3000 clones avant traitement et 60
après traitement.

Préparation de l'ADN (taille comprise entre 100 et 300kb).

Mise en bouts francs de l'ADN.

L'ADN est mis en bouts francs par élimination des séquences 3'
sortantes et remplissage des séquences 5' sortantes. Cette
opération est faite avec : enzyme de Klenow, T4 polymérase, les 4
20 nucléotides triphosphates.

Préparation des extrémités. Exemple HindIII

L'addition de l'ADN sur le vecteur se fait au moyen d'oligo-
nucléotides reconnaissant la séquence HindIII modifiée du vecteur.
25 Ils contiennent des sites de restriction rares pour permettre les
clonages ultérieurs (Swal ; NotI). cette technique est dérivée de celle
de : Elledge SJ, Mulligan JT, Ramer SW, Spottswood M, Davis RW.
Proc Natl Acad Sci U S A 1991 Mar 1;88(5):1731-5

Deux oligonucléotides complémentaires sont utilisés :

30 Oligo 1 : 5'-GCTTATTTAAATATTAATGCGGCCGCCCGGG-3'
(SEQ ID N°25)

Oligo 2 : 5'-CCCGGGCGGCCGCATTAATATTTAAATA-3' (SEQ ID
N°26)

Ils sont phosphorylés en 5' par la polynucléotide kinase de T4 en présence d'ATP, après leur hybridation. Cette étape de phosphorylation peut être éliminée en utilisant les oligo-nucléotides déjà phosphorylés.

5 La ligation de cet adaptateur double brin avec l'ADN à insérer dans un vecteur est faite par la ligase de T4 en présence d'un très grand excès d'adaptateur (1000 molécules d'adaptateur pour une molécule d'ADN à insérer), en 15 heures à 14°C.

10 L'excès d'adaptateur est éliminé par électrophorèse sur un gel d'agarose et les molécules d'intérêt sont récupérées du gel par hydrolyse de celui-ci par de l'agarase ou par électroélution.

Ligation vecteur- ADN .

15 La ligation se fait à 14°C sur 15 heures avec 10 molécules de vecteur pour une molécule d'insert.

Transformation.

20 La souche réceptrice est la souche DH10B. La transformation se fait par électroporation. Pour exprimer la résistance à la tétracycline, les transformants sont incubés à 37 °C pendant 1 heure en milieu sans antibiotique. La sélection des clones se fait par culture pendant une nuit , sur milieu gélosé LB additionné de tétracycline à 5µg.ml⁻¹.

Exemple 11 : CONJUGAISON CLONE A CLONE ENTRE E. COLI ET STREPTOMYCES

CONJUGAISON ENTRE E COLI SOUCHE S17.1 CONTENANT PPM803 ET STREPTOMYCES LIVIDANS TK 21

30

Introduction

Il est possible d'effectuer des conjugaisons entre *E. coli* et *Streptomyces* (Mazodier et al, 1989). L'adaptation de cette méthode en développant
35 une technique dite en goutte où l'on mélange 10 µl d'une culture de *E.*

coli contenant un vecteur recombinant à une goutte de *S. lividans* récepteur consiste à réaliser une transformation de clone à clone en s'assurant qu'à la fin de l'opération toute la banque construite dans *E. coli* est introduite dans *S. lividans*. Une transformation en vrac amènerait
5 obligatoirement à une multiplication des clones de *Streptomyces* transformants afin d'être pratiquement sûr que la banque dans *E. coli* est complètement représentée dans *S. lividans*.
De plus cette méthode est facilement automatisable.

10 **Essais préliminaires**

Conjugaison entre *E. coli* souche S17.1 contenant le vecteur pOSV303 et *S. lividans* TK21.

Dans ces conditions, on mélange 6×10^6 cellules de *E. coli* avec 2×10^6
15 spores pré-germées de *S. lividans* dans un volume final de 20 μ l.

Mise au point de la méthode

Il est connu que l'ADN extrait de certains actinomycètes est modifié et de
20 ce fait ne peut être introduit dans certaines souches de *E. coli* sans qu'il soit restreint. La souche de *E. coli* DH10B qui accepte ces ADN n'est pas capable de transférer à *Streptomyces* un plasmide ne contenant que *oriT*, et il est donc nécessaire d'en construire une. Il faudrait y introduire
par intégration dans le chromosome un dérivé de RP4 capable de fournir
25 en trans toutes les fonctions nécessaires pour assurer le transfert des clones recombinants contenant l'origine de transfert *oriT*.

Exemple 12 : Construction d'une banque cosmidique dans *E. coli* et *Streptomyces lividans* : Clonage de l'ADN du sol

30

L'objectif est la construction d'une librairie d'ADN de grande taille issue de l'environnement, sans étape préalable de culture des microorganismes, dans le but d'accéder aux gènes métaboliques de bactéries (ou de tout autre organisme) que l'on ne sait pas cultiver dans
35 des conditions standard de laboratoire.

La procédure décrite a été utilisée pour générer une banque d'ADN dans *Escherichia coli* utilisant le cosmide navette *E. coli*-*S. lividans* pOS700I et de l'ADN extrait et purifié de la fraction bactérienne d'un sol . Cette dernière méthode permet d'obtenir de l'ADN d'une grande pureté et d'une taille moyenne de 40 kb. Aussi, afin d'éviter pour le clonage une digestion partielle de l'ADN extrait, a été adoptée une stratégie alternative basée sur l'utilisation de l'enzyme terminale tranférase qui permet d'ajouter des queues de polynucléotides aux extrémités 3' de l'ADN et du vecteur.

5 µg d'ADN ont été extraits de 60 mg de sol de " la Côte Saint André " selon le protocole décrit à l'exemple 3 et traités avec de la terminale transférase (Pharmacia) pour rallonger les extrémités 3' avec un polynucléotide monotone (poly T) (Exemple 10).

Le cosmide intégratif pOS700I est préparé selon le protocole B1, Orsay. Après une étape classique de purification en présence de phénol/chloroforme, l'ADN et le vecteur sont assemblés en mélangeant une molécule de vecteur et une molécule d'ADN inséré. Le mélange est ensuite encapsidé dans les têtes de bactériophages lambda (kit Amersham) qui servent à transfecter *E. coli* DH10B. Les cellules transfectées sont ensuiteensemencées sur milieu LB agar en présence d'ampicilline pour sélection des recombinants résistants à cet antibiotique.

Une banque d'environ 5000 clones d'*E. coli* résistants à l'ampicilline a été obtenue. Chaque clone a été ensemencé en milieu LB ou TB + ampicilline dans un puits de microplaque (96 puits) et conservé à -80°C .

La séquence aux sites d'insertions des fragments du sol dans le vecteur, pOS700I, générés pendant la construction de la banque a été analysée. Pour cela 17 cosmides de la banques ont été purifié et séquencé avec une amorce, seq.5' CCGCGAATTCTCATGTTTGACCG 3', qui hybride entre les site BamHI et le site de clonage HindIII présente dans le vecteur.

Les séquences obtenues ont permis d'estimer que la longueur des queues homopolymériques aux points de jonctions est très variable, entre 13 et 60 poly-dA/dT. Au-delà des queues, les séquences des fragments du sol ainsi générées possèdent un pourcentage en G+C entre 53 et 70 %. Des pourcentages si élevés étaient inattendus, mais des résultats similaires ont été déjà reportés sur des préparations brut d'ADN à partir de sol (Chatzinotas A. *et al.*, 1998).

Une stratégie de "pooling" de 48 ou 96 clones a été utilisée pour l'analyse de la richesse microbienne et métabolique. L'ADN cosmique extrait à partir de ces "pools" de clones a été utilisé ensuite pour réaliser des expériences de PCR ou d'hybridation.

15

Exemple 13 : Diversité de l'ADN ribosomique 16S au sein de l'ADN cloné.

a) Matériels et méthodes

Les cosmides de la banque sont extraits à partir de pools de clones par lyse alcaline puis sont purifiés sur gradient de chlorure de césium, afin de prélever la bande d'ADN cosmique sous forme super-enroulée et dans le but d'éliminer tout ADN chromosomique d'*Escherichia coli* pouvant interférer dans l'étude.

Après linéarisation des cosmides par action de la nucléase S1 (50 unités, 30 minutes à 37°C), les séquences d'ADNr 16S contenues dans les pools de clones sont amplifiées dans les conditions standard d'amplification, à partir des amorces universelles 63f (5'-CAGGCCTAACACATGCAAGTC-3') et 1387r (5'-GGGCGGWGTGTACAAGGC-3') définies par MARCHESI *et al.* (1998). Les produits d'amplification d'environ 1.5 kilobases sont purifiés à partir du kit Qiaquick gel extraction (Qiagen) puis directement clonés dans le vecteur pCR II (Invitrogen) chez *Escherichia coli* TOP10, selon les instructions du fabricant. L'insert est alors amplifié à l'aide des amorces M13 Forward et M13 reverse spécifiques au site de clonage du vecteur

pCR II. Les produits d'amplification de taille attendue (environ 1,7 kb) sont analysés par RFLP (Restriction Fragment Length Polymorphism) à l'aide des enzymes CfoI, MspI et BstUI (0,1 unités) afin de sélectionner les clones à séquencer. Les profils de restriction obtenus sont séparés
5 sur gel d'agarose Metaphore 2.5% (FMC Products) contenant 0,4 mg de bromure d'éthidium par ml.

Les séquences d'ADNr 16S sont alors déterminées directement en utilisant les produits PCR purifiés par le kit " Qiaquick gel extraction " à l'aide des amorces de séquençage définies par Normand (1995). Les
10 analyses phylogénétiques sont obtenues en comparant les séquences avec les séquences d'ADNr 16S procaryotes rassemblées dans la base de données Ribosomal Database Project (RDP), version 7,0 MAIDAK et al.(1999) grâce au programme SIMILARITY MATCH, permettant d'obtenir les valeurs de similarité par rapport aux séquences de la base
15 de données.

b) Résultats

Pour déterminer la diversité phylogénétique représentée dans la banque, 47 séquences du gène ARNr16S ont été isolées à partir de
20 pools de 288 clones et ont été séquencées dans leur quasi-totalité. Les résultats sont rapportés dans le tableau 7.

L'analyse des séquences par interrogations des bases de données révèle que la majorité des séquences (>61%) présentent des
25 pourcentages de similarité inférieurs ou égaux à 95% avec des espèces bactériennes identifiées (tableau 7). Sur les 47 séquences analysées, 28 séquences ont pour plus proches voisins des bactéries non cultivées, dont les séquences ont été directement issues d'ADN extrait de l'environnement. La majorité de ces séquences présentent par ailleurs
30 des pourcentages de similarité très faibles (88-95%), 17 séquences sur 28 diffèrent ainsi de plus de 5% par rapport à leurs voisins les plus proches.

Parmi les séquences pouvant être classées dans un groupe phylétique, une majorité de séquences appartiennent à la sous classe a
35 des protéobactéries (18 séquences avec un pourcentage de similarité

compris entre 89 et 99%). Un second groupe de séquences est représenté par la sous classe g des protéobactéries, comprenant 9 séquences dont les pourcentages de similarité varient entre 84 et 99%). Les groupes des b-protéobactéries, d-protéobactéries, firmicutes à bas G+C% et à haut G+C% comprennent respectivement 1, 4, 3 et 5 séquences. Seule une séquence n'a pu être classée au sein des grands groupes taxonomiques bactériens définis : la séquence a22.1(19), son plus proche voisin *Aerothermobacter marianas* (avec une similarité de 89%) étant lui même une souche isolée de l'environnement marin et non classifiée à l'heure actuelle. Enfin, 6 séquences peuvent être classées au sein du groupe des *Acidobacterium/ Holophaga*. Ce groupe présente la particularité de n'être représenté que par deux bactéries cultivées *Acidobacterium capsulatum* et *Holophaga foetida*, l'ensemble de ce groupe étant composé par des bactéries dont seul le gène ARNr16S a été détecté par amplification et clonage à partir d'ADN extrait d'échantillon de l'environnement (principalement de sol), Ludwig *et al* (1997). Les faibles valeurs de similarité entre les différentes séquences composant ce groupe laisse présager une grande hétérogénéité et diversité au sein de ce groupe.

L'ensemble des résultats est représenté sur le tableau 7.

Ces résultats montrent que les séquences contenues dans la banque cosmidique proviendraient de micro-organismes non seulement diversifiés phylogénétiquement mais surtout de micro-organismes n'ayant jamais été isolés jusqu'à ce jour.

Les résultats du séquençage des ADN amplifiés ont permis d'établir un arbre phylogénétique des organismes présents dans l'échantillon de sol dont les séquences caractérisées sont originaires.

L'arbre phylogénétique représenté à la figure 7 a été réalisé à partir de l'alignement des séquences par le logiciel MASE (Faulner et Jurak, 1988) et corrigé par la méthode des 2 paramètres de Kimura (1980), et à l'aide de l'algorithme Neighbour Joining (Saitou et Nei 1987). L'analyse phylogénétique a permis de comparer les séquences ADNr 16S clonées dans la banque d'ADN du sol, avec les séquences d'ADNr 16S procaryotes rassemblées dans les bases de données Ribosomal

Database Project (RDP), (version 7.0, programme SIMILARITY-MATCH, Maidak *et al* 1999), et dans la base GenBank grâce au logiciel BLAST 2.0 (Atschul *et al*, 1997).

5

Exemple 14 : Présélection génétique de la banque pour l'évaluation de la richesse métabolique

10 Pour caractériser la banque obtenue en terme de diversité métabolique et identifier les clones contenant des inserts portant des gènes pouvant être impliqués dans des voies de biosynthèse, il a été développé selon l'invention des techniques de criblage génétique basées sur des méthodes PCR afin de détecter et d'identifier des gènes PKS de type I.

15

1 Souches bactériennes, plasmides et conditions de culture

20 *S. coelicolor* ATCC101478, *S. ambofaciens* NRRL2420, *S. lactamandurans* ATCC27382, *S. rimosus* ATCC109610, *B. Subtilis* ATCC6633 et *B. licheniformis* THE1856 (collection RPR) ont été utilisés comme sources d'ADN pour les expériences de PCR. *S. lividans* TK24 est la souche hôte utilisée pour le cosmide navette POSI700.

25 Pour la préparation d'ADN génomique, de suspensions de spores, de protoplastes et pour la transformation de *S. lividans*, on a suivi les protocoles standard décrits dans Hopwood *et al.* (1986).

Escherichia coli Top10 (INVITROGEN) a été utilisé comme hôte pour le clonage des produits PCR et *E. coli* Sure (STRATAGENE) a été utilisé comme hôte pour le cosmide navette pOS7001. Les conditions de culture de *E. coli*, la préparation de plasmides, la digestion de l'ADN, 30 l'électrophorèse sur gel d'agarose ont été réalisées suivant les procédures standard (Sambroock *et al.*, 1996).

2. Amorces PCR:

35 Les couples d'amorces a1-a2 et b1-b2 ont été définis par l'équipe de N. Bamas-Jacques et leur utilisation a été optimisée pour le

criblage de l'ADN des souches pures et de la banque du sol pour la recherche de gènes codant PKSI)

Tableau 8 :

5 **Amorces PCR homologues aux gènes PKSI utilisées pour le criblage de la banque.**

a1 (+)	5' CCSCAGSAGCGCSTSTTSCTSGA 3'
a2 (-)	5' GTSCCSGTSCCGTGSGTSTCSA 3'
b1	5' CCSCAGSAGCGCSTSCTSCTSGA 3'
b2	5' GTSCCSGTSCCGTGSGCCTCSA 3'

10 **Conditions d'amplification :**

Pour la recherche de PKS I à partir de l'ADN de souches pures, le mélange d'amplification contenait : dans un volume final de 50 µl, entre 50 et 150 ng d'ADN génomique, 200 µM de dNTP, 5mM de MgCl₂ final, 7% de DMSO, tampon 1x Appligene, 0,4 µM de chaque primer et
 15 2,5U de Taq Polymerase Appligène. Les conditions d'amplification utilisées sont : dénaturation à 95°C pendant 2 minutes, hybridation à 65°C pendant 1 minute, élongation à 72°C pendant 1 minute, pour le premier cycle, suivi par 30 cycles où la température est diminuée
 20 jusqu'à 58°C comme décrit dans K. Seow *et al.*, 1997. L'étape d'extension finale s'effectue à 72°C pendant 10 minutes.

Pour la recherche de PKS I à partir de l'ADN de la banque, les conditions PCR sont les mêmes que ci-dessus pour le couple a1-a2 en
 25 utilisant entre 100 et 500 ng de cosmide extrait de pools de 48 clones. Pour le couple d'amorces b1-b2, 500ng de cosmides issus de pools de 96 clones ont été utilisés. Le mélange d'amplification contenait 200 µM

de dNTP, 2,5mM de MgCl₂ final, 7% de DMSO, tampon 1x Quiagen, 0,4 µM de chaque primer et 2,5U de Taq polymerase Hot-start (Qiagen). Les conditions d'amplification utilisées sont : dénaturation 15' à 95°C suivie par 30 cycles : 1' de dénaturation à 95°C + 1' d'hybridation à 65°C pour le premier cycle et 62°C pour les autres cycles, 1' d'élongation à 72°C, étape d'extension finale de 10' à 72°C.

L'identification des clones positifs à partir des pools de 48 ou 96 clones est effectuée à partir des répliques des microplaques mères correspondantes sur milieu solide ou toute autre méthode standard de réplication.

3 Sous-clonage et séquençage

Les produits PCR des clones identifiés ont été séquencés selon le protocole suivant :

Les fragments sont purifiés sur gel d'agarose (Gel Extraction Kit (Qiagen)) et clonés dans *E.coli* TOP 10 (Invitrogen) à l'aide du kit TOPO TA cloning kit (Invitrogen). L'ADN plasmidique de sous-clones est extrait par lyse alcaline sur un Biorobot (Qiagen) et dialysé durant 2 h sur membrane VS 0,025µm (Millipore). Les échantillons sont séquencés avec les amorces M13 " Universal " et " Reverse " sur le séquenceur ABI 377 96(PERKIN ELMER).

4) Résultats

Définition et validation des amorces PCR

Deux régions très conservées de PKS du type I d'actinomycètes, comprenant le site actif de l'enzyme, ont été ciblées pour l'amplification de gènes homologues avec des amorces dégénérées. Ces deux régions correspondent aux séquences PQQR(L)(L)LE et VE(A)HGTGT respectivement.

Des amorces (tableau 8) ont été testées avec l'ADN de souches productrices ou non de macrolides: *Streptomyces coelicolor*,

Streptomyces ambofaciens, producteur de spiramycine, et *Saccharopolyspora erythraea*, producteur de l'erythromycine. Quelles que soient les amorces utilisées, des bandes représentant des fragments d'environ 700 pb et correspondant à la longueur du fragment attendu, ont été obtenues avec toutes les souches.

Ces résultats démontrent la spécificité des amorces a et b pour les gènes PKS I de souches productrices ou de gènes silencieux chez *S. coelicolor*.

Le séquençage des produits PCR obtenus avec le couple d'amorces a1-a2 a permis d'identifier, à partir de la souche *S. ambofaciens*, la séquence d'un gène KS déjà décrite (Demande de brevet européen n° EP0791656) comme appartenant à la voie de biosynthèse du planténolide, précurseur macrolidique de la spiramycine, et deux séquences jamais décrites, Stramb 9 et Stramb12, (voir liste séquences).

En ce qui concerne, *S. erythraea*, la méthode de criblage a permis l'identification d'une séquence de KS (sacery17) identique à celle du KS du module 1 déjà publiée dans Genbank (Numéro d'accès M63677), codant pour la synthétase 1 (DEBS1) du 6-deoxyérythronolide B. Une autre séquence non corrélée à la voie de biosynthèse de l'érythromycine a été identifiée et il s'agit de la séquence SEQ ID N° 32.

Conclusion

Une méthode pour analyser la présence de gènes codant pour les PKS du type I par PCR à partir de différents micro-organismes a été mise au point. La structure très conservée du domaine de la keto-synthétase du type I a permis de réaliser une méthode PCR basée sur l'utilisation d'amorces dégénérées biaisées en GC pour le choix des codons.

Cette approche montre la possibilité d'identifier des gènes ou clusters impliqués dans la voie de biosynthèse des polyketides du type I. Le clonage de ces gènes permet la création d'une collection qui pourra

ensuite être utilisés pour construire des hybrides polyketides. Le même principe peut être appliqué à d'autres classes d'antibiotiques.

Les résultats obtenus ici montrent aussi la présence de gènes pouvant appartenir à des clusters silencieux (SEQ ID N° 30 à 32).

- 5 La présence de clusters silencieux a été déjà documentée dans *S. lividans* et leurs expressions sont déclenchées par des régulateurs spécifiques ou pleiotropiques (Horinouchi et al. ;Umeyama et al. 1996) . Ces résultats suggèrent que la détection de gènes appartenant à des voies dites silencieuses codent en réalité pour des enzymes actives
- 10 capable de diriger, en association avec les autres enzymes spécifiques de la voie, les étapes enzymatiques nécessaires pour la synthèse des métabolites secondaires.

Criblage de la banque

15

Le criblage a été effectué dans les conditions décrites dans la section Matériels et Méthodes en utilisant les couples d'amorces validées à partir de souches productrices.

- En présence du couple d'amorces a1-a2 , la taille des produits
- 20 PCR obtenus à partir de l'ADN cosmidique extrait de pools de 48 ou 96 clones était d'environ 700 bp, donc en accord avec les résultats attendus.

L'intensité des bandes obtenues était variable, mais une seule bande d'amplification était présente pour chaque pool d'ADN cible.

- 25 Dans ces conditions, 8 groupes d'ADN cible ont été détectés, correspondant à 9 clones positifs après déréplication.

Le criblage effectué avec le second couple d'amorces, b1-b2, a donné des résultats d'amplification moins spécifiques puisque de nombreuses bandes satellites étaient observées à côté de la bande de 700 bp.

- 30 Néanmoins, 9 groupes d'ADN cible ont été détectés, correspondant à 14 clones positifs après déréplication à partir de ces clones positifs, l'ADN a été extrait pour les étapes de séquençage et de transformation de *S. lividans*.

Analyse des cosmides

La digestion des cosmides identifiés par PCR avec l'enzyme DraI, reconnaissant un site riche en AT, libère un fragment supérieur à 23 kb (figure 22). Ceci suggère que la méthode PCR cible préférentiellement l'ADN du sol contenant un haut pourcentage en G+C. Ce résultat est la conséquence de la dégénérescence des amorces utilisées, biaisées en GC pour le choix des codons. Les inserts, comme attendu dans le cas de cosmides, ont une taille supérieure à 23 kb, sauf dans un cas (clone a9B12), ce qui pourrait traduire une certaine instabilité des cosmides. D'autre part, parmi tous les clones sélectionnés, seulement deux d'entre eux, GS.F1 et GS.G11, ont montré le même profil de restriction indiquant un faible taux de redondance dans la banque.

Les cosmides sélectionnés ont été transférés dans *Streptomyces lividans* par transformation de protoplastes en présence de PEG 1000. L'efficacité de transformation varie entre 30 et 1000 transformants par µg d'ADN cosmidique utilisé.

Séquençage et analyse phylogénétique des gènes PKS I du sol

La méthode de PCR mise à point sur les souches pures a été utilisée comme décrite sur les cosmides de la banque et 24 clones ont ainsi été identifiés.

Les produits de PCR d'environ 700 bp obtenus à partir de l'ADN de deux pools (48 clones) et de 8 clones uniques, ont été clonés, après purification sur gel d'agarose, et séquencés. Cela a permis l'identification de 11 séquences.

L'alignement des séquences protéiques déduites PKSs I du sol avec d'autres PKSs I présentes dans différents micro-organismes (figure 24) montre la présence d'une région très conservée qui correspond à la région consensus du site active de la b-kétoacyl synthétase.

L'analyse des séquences obtenues avec la méthode "Codonpreference" (Gribskov *et al.*, 1984 ; Bibb *et al.*, 1984) a révélé la présence d'un fort biais dans l'usage des codons riches en G+C dans une seule phase de lecture. Les protéines déduites selon cette phase
5 de lecture montrent une forte similarité avec des KSs du type I connus (programme Blast). En particulier, la similarité entre les séquences de KSs du sol et des KSs du cluster de l'érythromycine est d'environ 53%.

Après déréplication d'un pool et identification du clone unique, la séquence du produit PCR obtenu à partir de ce clone est identique à
10 celle du pool ce qui confirme la fiabilité de la méthode utilisée.

L'analyse de la séquence du produit PCR d'un clone a permis l'identification probable de 3 gènes KSI différents. Une de ces séquences (SEQ ID N° 34) a une similarité de 98,7% avec la séquence d'un autre pool, suggérant qu'elles codent pour la même enzyme. Les
15 deux autres séquences sont différentes mais fortement homologues.

Ici, il est décrit pour la première fois le clonage et l'identification dans une banque d'ADN du sol de voies de biosynthèse de métabolites secondaires contenant des gènes codant des KS du type I.

Le pourcentage élevé en G+C des séquences du sol suggère
20 qu'elles puissent dériver de génomes ayant un usage des codons similaire à ceux d'actinomycètes.

Même si les données disponibles dans la littérature sont réduites, on sait que les gènes codant des PKS du type I sont très diversifiés de par leur organisation physique dans le génome, la taille et le nombre de
25 modules contenus dans chaque gène.

La présence de plusieurs domaines provenant d'un seul clone est une confirmation de leur appartenance à des clusters de polyketides asymétriques. Dans un seul cas, deux clones semblent former un contigue puisqu'ils partagent la même séquence pour le domaine KS.

30 La taille des régions génétiques impliquées dans la synthèse des PKS I varie entre quelques kb pour la pénicilline à environ 120 kb pour la rapamycine. La dimension des inserts cosmidiques peut donc ne pas être suffisante pour l'expression des clusters les plus complexes.

Des gènes codant pour des PKSs I, capables de travailler de
35 façon itérative comme les PKS II et de contrôler la synthèse de

polykétides aromatiques, ont été décrits (Jae-Hyuk et al., 1995). L'étude des clusters des PKSs I du sol pourrait apporter encore des nouveautés dans ce domaine.

5 **5. Identification de 6 gènes codant des polykétides synthases .**

On poursuivant le criblage de la banque de cosmides selon les protocoles décrits dans le présent exemple, les inventeurs ont identifiés
10 un clone de cosmide contenant un insert de 34071 pb contenant plusieurs cadres ouverts de lecture codant pour des polypeptides du type polykétide synthase.

Plus précisément, le cosmide ainsi identifié par criblage de la banque contient six cadres ouverts de lecture codant pour des
15 polypeptides polykétide synthase ou pour des polypeptides fortement apparentés, des peptides synthase non ribosomiques. Une carte détaillée de ce cosmide est représentée à la figure 36.

La séquence nucléotidique complète du cosmide constitue la séquence SEQ ID N°113 du listage de séquences. L'insert d'ADN
20 contenu dans la séquence SEQ ID N°113 constitue la séquence nucléotidique complémentaire (brin -) de la séquence nucléotidique codant pour les différents polykétides synthases.

La séquence nucléotidique de l'insert d'ADN contenue dans le cosmide de la figure 36 qui comprend les cadres de lecture ouverts
25 codant pour les polypeptides polykétides synthases (brin +) est schématisée sur la figure 37 et constitue la séquence SEQ ID N°114 du listage de séquences.

De plus, une carte détaillée des différents cadres de lecture ouverts contenus dans l'insert d'ADN de ce cosmide est représentée à la
30 figure 37.

Les caractéristiques des séquences nucléotidiques comprenant des cadres ouverts de lecture contenus dans l'insert d'ADN de ce cosmide sont détaillées ci-après.

Séquence ORF1

La séquence orf1 comprend un cadre ouvert de lecture partielle d'une longueur de 4615 nucléotides. Cette séquence constitue la
5 séquence SEQ ID N°115, qui débute au nucléotide en position 1 et se termine au nucléotide en position 4615 de la séquence SEQ ID N°114.

La séquence SEQ ID N°115 code pour le polypeptide ORF1 de 1537 acides aminés, ce polypeptide constituant la séquence SEQ ID N°121.

10 Le polypeptide de séquence SEQ ID N°121 est apparenté aux peptides synthases non ribosomiques. Ce polypeptide possède un degré d'identité en acides aminés de 37% avec le peptide synthase de *Anabaena* sp.90 référencé sous le numéro d'accès « emb CACO1604.1 » dans la base de données Genbank.

15

Séquence ORF2

La séquence nucléotidique orf2 a une longueur de 8301 nucléotides et constitue la séquence SEQ ID N°116, qui débute au
20 nucléotide en position 4633 et se termine au nucléotide en position 12933 de la séquence SEQ ID N°114.

La séquence ORF2 code pour le peptide ORF2 d'une longueur de 2766 acides aminés, ce polypeptide constituant la séquence SEQ ID N°122.

25 Le polypeptide de séquence SEQ ID N°122 possède une identité de séquence en acides aminés de 41% avec la séquence MtaD de *Stigmatella aurantiaca* référencée sous le numéro d'accès « gb AAF 19812.1 » de la base de données GENBANK.

Le polypeptide ORF2 constitue une polykétide synthase.

30

Séquence ORF3

La séquence nucléotidique orf3 a une longueur de 5292 nucléotides et constitue la séquence SEQ ID N°117. La séquence SEQ
35 ID N°117 correspond à la séquence qui débute au nucléotide en position

12936 et qui se termine au nucléotide en position 18227 de la séquence SEQ ID N°114.

La séquence nucléotidique SEQ ID N°117 code pour le polypeptide polykétide synthase ORF3 de 1763 acides aminés, ce
5 polypeptide constituant la séquence SEQ ID N°123 selon l'invention.

Le polypeptide ORF3 de séquence SEQ ID N°123 possède une identité de 42% en acides aminés avec la séquence MtaB de *Stigmatella aurantiaca* référencée sous le n° d'accès « gb AAF 19810.1 » de la base de données GENBANK.

10

Séquence ORF4

La séquence nucléotidique orf4 a une longueur de 6462 nucléotides et constitue la séquence SEQ ID N°118 selon l'invention.

15 La séquence nucléotidique SEQ ID N°118 correspond à la séquence débutant au nucléotide en position 18224 et se terminant au nucléotide en position 24685 de la séquence nucléotidique SEQ ID N°114.

La séquence nucléotidique SEQ ID N°118 code pour le polypeptide polykétide synthase ORF4 de 2153 acides aminés, ce
20 polypeptide constituant la séquence SEQ ID N°124 selon l'invention.

Le polypeptide ORF4 de séquence SEQ ID N°124 possède une identité de séquence en acides aminés de 46% avec la séquence epoD de *Sorangium cellulosum* référencée sous le n° d'accès « gb
25 AAF62883.1 de la base de données GENBANK.

Séquence ORF5

La séquence nucléotidique orf5 a une longueur de 5088
30 nucléotides et constitue la séquence SEQ ID N°119 selon l'invention.

La séquence SEQ ID N°119 correspond à la séquence débutant au nucléotide en position 24682 et se terminant au nucléotide en position 29769 de la séquence nucléotidique SEQ ID N°114.

La séquence nucléotidique SEQ ID N°119 code pour le polypeptide polykétide synthase ORF5 de 1695 acides aminés, ce polypeptide constituant la séquence SEQ ID N°125 selon l'invention.

Le polypeptide polykétide synthase ORF5 de séquence SEQ ID N°125 possède une identité en acides aminés de 43% avec la séquence epod de *Sorangium cellulosium* référencé sous le n° d'accès « gb AAF 62883.1 » de la base de données GENBANK.

Séquence ORF6

10

La séquence nucléotidique orf6 a une longueur de 4306 nucléotides, et constitue la séquence SEQ ID N°120 selon l'invention. La séquence nucléotidique SEQ ID N°120 correspond à la séquence débutant au nucléotide en position 29766 et se terminant au nucléotide en position 34071 de la séquence SEQ ID N°114.

La séquence SEQ ID N°120 contient un cadre ouvert de lecture partielle codant pour le polypeptide ORF6 de 1434 acides aminés du type polykétide synthase, ce polypeptide constituant la séquence SEQ ID N°126 selon l'invention.

Le polypeptide de séquence SEQ ID N°126 possède une identité en acides aminés de 43% avec la séquence epod de *Sorangium cellulosum* référencée sous le numéro d'accès « gb AAF 62883.1 » de la base de données GENBANK.

EXEMPLE 15: Construction de vecteurs navettes de type BAC intégratifs chez *Streptomyces*

Construction de vecteurs navettes du type BAC intégratifs et conjugatifs chez *Streptomyces*

30

15.1 Construction du vecteur pMBD-1

Le vecteur BAC pMBD-1 a été obtenu selon les étapes suivantes:

Etape 1: Le vecteur pOSVO10 a été soumis à une digestion par les enzymes PsTI et BstZ17I afin d'obtenir un fragment nucléotidique de 6,3 kb.

5 **Etape 2:** Le vecteur pDNR-1 a été digéré par les enzymes PstI et PvuII afin d'obtenir un fragment nucléotidique de 4,145 kb.

10 **Etape 3:** Le fragment nucléotidique de 6,3 kb provenant du vecteur pOSV017 a été fusionné par ligation au fragment de 4,15 kb provenant du vecteur pDNR-1, afin de produire le vecteur pMBD-1, comme cela est illustré à la figure 30.

15.2 Construction du vecteur pMBD-2

15 Le vecteur pMBD-2 est un vecteur du type BAC contenant une boîte intégrative « ϕ c31 int- Ω hyg ».

20 ϕ c31 est un phage tempéré à spectre d'hôte large dont le site d'attachement (attP) est bien localisé. Le fragment ϕ c31 int est le fragment minimal de l'actinophage ϕ c31 capable d'induire l'intégration d'un plasmide dans le chromosome de *Streptomyces Lividans*.

Ω hyg est un dérivé de l'interposon Ω capable de conférer la résistance à l'hygromicine chez *E.coli* et *S.Lividans*.

25 Des vecteurs BAC contenant le système d'intégration ϕ c31 sont décrits par SOSIO et al. (2000) et dans la demande PCT n°99 6734 publiée le 29 Décembre 1999.

Le vecteur BAC pmBD-2 a été construit selon les étapes suivantes:

Etape 1: Construction d'une boîte intégrative ϕ c31int Ω hyg dans un plasmide multicopies de *E.coli*.

30 On a tout d'abord amplifié le fragment ϕ c31int à partir du plasmide pOJ436 à l'aide du couple d'amorces suivant:

35 - L'amorce EV ϕ c31I (SEQ ID N°109) (qui permet d'introduire un site EcoRV à l'extrémité 5' de la séquence ϕ c31) et l'amorce BII ϕ c31F (SEQ ID N°110) (qui permet l'introduction d'un site BglII à l'extrémité 3' de la séquence ϕ c31).

Le fragment Ω hyg a été obtenu par digestion à l'aide de l'enzyme BamHI du plasmide pHP45 Ω hyg décrit par BLONDELET-ROUAULT (1997).

Puis la boîte intégrative ϕ c31 int- Ω hyg a été clonée dans le vecteur pMCS5 digéré par les enzymes BglII et EcoRV.

Etape 2: Construction du vecteur pMBD-2.

Le chromosome artificiel bactérien pBAce3.6 décrit par FRENGEN et al. (1999) a été digéré par l'enzyme NheI puis traité par l'enzyme Eco polymérase.

Puis, le vecteur pMCS5 ϕ c31 int- Ω hyg a été digéré par les enzymes SnaBI et EcoRV afin de récupérer la boîte intégrative.

La carte détaillée du vecteur pMBD2 est représentée à la figure 31.

15.3 Construction du vecteur pMBD-3.

Le vecteur pMBD-3 est un vecteur intégratif (ϕ c31 int) et conjuguatif (OriT) du type BAC, qui comprend le marqueur de sélection Ω hyg.

La carte du vecteur pMBD-3 ainsi que son procédé de construction sont illustrés à la figure 31.

Le vecteur pMBD-3 a été obtenu en amplifiant le gène OriT à partir du plasmide pOJ436 à l'aide du couple d'amorces de séquences SEQ ID N° 111 et SEQ ID N°112 qui contiennent des sites de restriction pacl.

Le fragment nucléotidique amplifié à l'aide des amorces SEQ ID N°111 et SEQ ID N°112 a été cloné dans le vecteur pMBD2 préalablement digéré par l'enzyme PacI. Le schéma de construction du vecteur pMBD-3 est illustré à la figure 31.

15.4 Construction du vecteur pMBD-4

La carte détaillée du vecteur pMBD-4 est représentée à la figure 32.

- 5 Le vecteur pMBD4 a été obtenu en clonant la boîte intégrative ϕ c31 int- Ω hyg dans le vecteur pCYTAC2.

15.5 Construction du vecteur pMBD-5

- 10 Le schéma de construction du vecteur pMBD-5 est illustré à la figure 33.

- Le vecteur pMBD-5 a été construit par recombinaison du fragment nucléotidique compris entre les deux sites loxP du vecteur pMBD-1 illustré à la figure 33 avec le site loxp contenu dans le vecteur
15 BAC désigné pBTP3, une carte détaillée du plasmide pBTP3 étant représentée à la figure 34.

15.6 Construction du vecteur pMBD-6

- 20 Le vecteur pMBD-6 a été construit en recombinant le fragment nucléotidique compris entre les deux sites loxP du vecteur pMBD-1 au niveau du site loxP du vecteur BAC pBeloBac11, comme représenté sur la figure 35.

TABEAU 1
Localisation des prélèvements d'échantillon et caractéristiques des sols utilisés
dans les différentes expériences. Les comptes microbiens directs en utilisant
la coloration à l'acridine orange ont été réalisés avant et après broyage du sol

Numéro	Origine	Texture	Quantité (%) de sable Limon Argile	Matière organique (g/kg de sol sec)	pH	Nombre de cellules avant broyage ^a (x10 ⁹ /g poids sol sec)	Nombre de cellules après broyage ^a (x10 ⁹ /g poids sol sec)
1	Australie	Argile sablonneuse	62	22	16	49,7	2,9(1,3)
2	Peyrat le Châ- teau, France	Argile sablonneuse	61	26	13	48,2	5,4(0,8)
3	Côte St-André, France	Terreau sablonneux	50	41	9	40,6	7,5(1,4)
4	Chazay d'Azergue, France	Terreau sablonneux argileux	34	47	19	13,9	4,2(0,6)
5	Guadeloupe, France	Argile	27	26	47	17,0	0,5(0,1)
6	Dombes, France	Terreau sablonneux Argileux	20	67	13	30,3	5,6(0,9)

^a n=3; déviation standard entre parenthèses.

TABLEAU 2
Amorces et sondes utilisées pour l'amplification PCR et
l'hybridation sur tâche

Amorce ou sonde	Cible ^{a)}	Séquence (5' à 3')	Référence n°
FGPS431 sonde	Universelle (1392-1406)	ACGGGCGGTGTGT(A/G)C	Amann et al., 1995
FGPS122 amorce	Bactéries (6-27)	GGAGAGTTTGATCATGGCTCAG	Amann et al., 1995
FGPS350 amorce	<i>Streptosporangium</i> (616-635)	CCTGGAGTTAAGCCCCCAAGC	Cette étude
FGPS643 sonde	<i>Streptosporangium</i> (122-142)	GTGAGTAACCTGCCCC(T/C)GACT	Cette étude
R499 amorce	<i>Bacillus anthracis</i>	TTAATTCACCTTGCAACTGATGGG	Patra et al., 1996
R500 amorce	<i>Bacillus anthracis</i>	AACGATAGCTCCTACATTTGGAG	Patra et al., 1996
C501 sonde	<i>Bacillus anthracis</i>	TTGCTGATACGGTATAGAACCTGGC	Patra et al., 1996
FGPS516 amorce	<i>S. lividans</i> OS48.3	TCCAGATCCTTGACCCGCAG	Cette étude
FGPS517 amorce	<i>S. lividans</i> OS48.3	CACGACATTGCACCTCCACCG	Cette étude
FGPS518 sonde	<i>S. lividans</i> OS48.3	CCGTGAGCCCGGATCAG	Cette étude

^{a)} Les positions sur le gène de l'ARNr 16S de *E. coli* sont données entre parenthèses. Pour *B. anthracis* et *S. lividans*, les amorces et les sondes ciblent des séquences chromosomiques spécifiques des organismes respectifs. Ces séquences ne sont pas localisées dans le gène de l'ARNr 16S. La cassette contenant la région cible de *S. lividans* est décrite par Clerc-Bardin et al. (non publié).

TABLEAU 3
Quantité d'ADN extrait à partir de différents sols après des traitements
de lyse selon les protocoles n°1 à 5 (µg ADN/g de poids de sol sec ± déviation standard)^a
Sols1, 2, 3 et 6; n= 3; sol 4: n=1.

Sol	Protocole de lyse numéro ^b				
	1	2	3	4a	4b 5a 5b
1. Australie	17+/-2	52+/-2	32+/-5	16 +/-3	33+/-2 59+/-1 27+/-0
2. Peyrat	29+/-2	58+/-1	40+/-2	29+/-2	18+/-3 56+/-1 15+/-1
3. Côte St-André	36+/-7	60+/-6	148+/-10	94+/-7	38+/-6 73+/-5 47+/-6
4. Chazay	9	16	ND	32	15 15 70
6. Dombes	4+/-2	26+/-3	43+/-1	61+/-	66+/-1 160+/-7 102+/-5

^a Quantification par imagerie de phosphorescence après hybridation sur tâche avec la sonde universelle FGPS431 (tableau 2).

^b1: Aucun traitement; 2: broyage à sec du sol (G); 3: Cr + homogénéisation Ultraturax (H);

4a: G + H + sonication Microtip (MT); 4b: G + H+ sonication Cup Hom (CH); 5a: Cr + H + NT + lyse chimique/ enzymatique. Voir aussi figure 1.

^c ND = Non déterminé.

Tableau 4 :
Amorces et sondes utilisées dans la caractérisation
moléculaire des ADN extraits du sol

	Cible (amorce ou sonde)	Séquence (5' - 3')	Position ^a
FGPS 612	Eubactéries (amorce)	C(C/T)AACT(T/C/A)CGTGCCAGCAGCC	506 - 525
FGPS 669	Eubactéries (amorce)	GACGTC(A/G)TCCCC(A/C)CCTTCCTC	1174 - 1194
FGPS 618	Eubactéries (sonde)	ATGG(T/C)TGTCGTCAGCTCG	1056 - 1073
FGPS 614	a-Protéobactéries (sonde)	GTGTAGAGGTGAAATTCGTAG	683 - 703
FGPS 615	b-Protéobactéries (sonde)	CGGTGGATGATGTGGATT	939 - 956
FGPS 616	g-Protéobactéries (sonde)	AGGTTAAACTCAAATGA	900 - 917
FGPS 621	Gram plus à bas GC% (sonde)	ATACGTAGGTGGCAAGCG	532 - 549
FGPS 617	Actinomycètes (sonde)	GCCGGGGTCAACTCGGAGG	1159 - 1149
FGPS 680	Streptomycètes (sonde)	TGAGTCCCCA(A/C/T)C(T/A)CCCCG	1132 - 1149
FGPS 619	Streptosporangium (sonde)	GCTTGGGGCTTAACTCCAGG	609 - 628

^a : position sur le gène ARNr16S d'*Escherichia coli*

Tableau 5 :
Efficacités d'extraction des cellules bactériennes sur gradient de Nycodenz et quantités d'ADN extrait.
Effet de l'incubation de l'échantillon de sol dans une solution d'extrait de levure 6%,
préalablement à la dispersion et à la centrifugation sur gradient de densité.

	Bactéries extraites	ADN extrait		Actinomycètes cultivables ^c cfu /g sol sec	Lyse directe ^d ng ADN/ g sol sec	Lyse en bloc d'agarose ^{d,e} ng ADN/ g sol sec
	Microflore totale ^a bactéries /g sol sec	Microflore cultivable ^b cfu /g sol sec				
Sans incubation						
Suspension de sol	1.3 10 ⁹ (± 0.1)	6.9 10 ⁶ (± 0.2)		8.6 10 ⁸ (± 1.2)		
Extrait cellulaire	1.9 10 ⁹ (± 0.2)	4.1 10 ⁶ (± 1.5)		2.5 10 ⁸ (± 0.7)	333 (± 35)	221 (± 70)
Efficacité d'extraction	15%	59%		38%		
Avec incubation dans extrait de levure 6%						
Suspension de sol	1.2 10 ⁹ (± 0.1)	7.6 10 ⁷ (± 1.1)		6.6 10 ⁷ (± 0.4)		
Extrait cellulaire	1.6 10 ⁹ (± 0.3)	5.3 10 ⁸ (± 1.4)		3.7 10 ⁹ (± 0.7)	344 (± 30)	341 (± 67)
Efficacité d'extraction	13%	7%		5%		

5 a : Dénombrement microscopique après coloration à l'acridine orange

b : Dénombrement sur milieu solide Trypcase-Soja 10%

c : Dénombrement sur milieu solide HV Agar, après enrichissement 20 minutes à 40°C dans une solution d'extrait de levure 6% - SDS 0,05%.

d : La quantité d'ADN extrait a été évaluée sur gel d'électrophorèse par rapport à une gamme étalon d'ADN de thymus de veau.

e : La quantification a été réalisée après digestion de l'agarose par action d'une b-agarase

Tableau 6 :
Caractérisation des ADN extraits en fonction de leur proportion en a, b, et g
sous classes des Protéobactéries, en Gram + à bas GC% et en Actinomycètes ; le signal d'hybridation avec la
sonde procaryote servant de référence 100%.

	a- Protéobactérie s	b- Protéobactérie s	g- Protéobactérie s	Gram+ bas GC%.	Actinomycètes	Streptomycètes
Extraction directe ^a	7.7 % (± 1.4)	5.3 % (± 0.5)	3.3 % (± 0.9)	3.1 % (± 1.7)	14.7 % (± 0.6)	0.8 % (± 0.1)
Extraction indirecte						
Lyse + CsCl	10.9 % (± 1.4)	6.4 % (± 1.4)	14.3 % (± 1.4)	7.9 % (± 1.4)	8.5 % (± 1.4)	3.0 % (± 1.4)
Lyse en bloc	2.9 % (± 1.4)	5.4 % (± 1.4)	11.1 % (± 1.4)	8.0 % (± 1.4)	11.3 % (± 1.4)	2.6 % (± 1.4)
Lyse en bloc + incubation YE	6.3 % (± 1.4)	7.5 % (± 1.4)	17.0 % (± 1.4)	18.1 % (± 1.4)	19.4 % (± 1.4)	4.6 % (± 1.4)

a : broyage dans un broyeur à billes de tungstène, à force centrifuge (protocole d'extraction décrit dans article Frostegard et al.)
YE : solution d'extrait de levure à 6%

Tableau 7:
Diversité des séquences d'ADNr16S contenues dans la banque cosmétique

N° pool (clone n°)	Voisin identifié le plus proche	% de similarité	Voisin le plus proche (classification, référence)	% de similarité
α -Protéobactéries				
a24.1 (2)	Azospirillum brasilense	97.7%		
a4-a6-a7 (7)	Azospirillum brasilense	95.4%		
a4-a6-a7 (23)	Azospirillum brasilense	88.9%	Str L-87 (a-proteobactérie) ¹	89.8%
a52-a53-a5 (15)	Azospirillum lipoferum	97.6%		
a49-a50-a51 (22)	Agrobacterium tumefaciens	95.0%	Clone JN15d (non publié)	95.5%
a49-a50-a51 (11)	Rhizobium sp	99.7%		
a4-a6-a7 (14)	Rhizobium sp	99.7%		
a30-a31-a32 (7)	Bradyrhizobium japonicum	99.4%		
a19-a20-a26 (5)	Bradyrhizobium genosp	93.3%	Clone DA122 (non publié)	95.9%
a37-a38-a39 (6)	Mesorhizobium sp.	98.9 %		
a19-a20-a26 (9)	Bradyrhizobium sp	90.2%	CloneS-26(aProtéobactérie) ²	95.9%
a46-a47-a48 (14)	Phyllobacterium rubiacearum	97.6 %		

TABLEAU 7 (suite 1)
Diversité des séquences d'ADNr16S contenues dans
la banque cosmétique

N° pool (clone n°)	Voisin identifié le plus proche	% de similarité	Voisin le plus proche (classification, référence)	% de similarité
a49-a50-a51 (1)	Caulobacter henricii	97.0%		
a1-a2-a3 (13)	Caulobacter sp.	96.3%		
a52-a53-a5 (8)	Mesorhizobium mediterraneum	92.1%	Clone DA122 (non publié)	94.8%
a34-a35-a36 (3)	Rhodobium orientis	91.8%	Clone (non publié)	95.1%
a1-a2-a3 (4)	Sphingomonas sp.	94.7%	Clone PAD23 (non publié)	95.1%
a8-a9-a10 (13)	Sphingomonas sp.	94.0%		
γ -Protéobactéries				
a40-a41-a42 (13)	Pseudomonas sp	98.9%	clone G26(g-Protéobactérie) ³	99.7%
a15-a16-a17 (12)	Lysobacter antibioticus	94.4%	clone vadinHA77(g-Protéo) ⁴	93.6%
a15-a16-a17 (5)	Xanthomonas sp	93.4%	clone vadinHA77(g-Protéo) ⁴	94.6%
a19-a20-a26 (13)	Luteimonas mephitis	92.9%	Souche rJ15 (non publié)	93.5%
a46-a47-a48 (6)	Methylobacter whittenburyi	88.3%	soil clone S-43(g-Protéo) ²	88.9 %
a11-a12-a13 (11)	Methylobacter whittenburyi	88.3%	soil clone S-43(g-Protéo) ²	88.9%

TABLEAU 7 (suite 2)
Diversité des séquences d'ADNr16S contenues dans
la banque cosmétique

N° pool (clone n°)	Voisin identifié le plus proche	% de similarité	Voisin le plus proche (classification, référence)	% de similarité
a34-a35-a36 (5)	Methylococcus capsulatus	84.9%	soil clone S-12 (d-Protéo) ²	85.6%
a43-a44-a45 (10)	Legionella birninghamensis	88.9%		
a8-a9-a10 (2)	Lamprocystis roseopersicina	87.5%	Clone 2-100C14 (non publié)	95.1%
β -Protéobactéries				
a27-a28-a29 (5)	Rhodocyclus tenuis	90.2%	Clone OPB37 (b-protéo) ⁵	91%
δ -Protéobactéries				
a8-a9-a10 (18)	Nannocystis exedens	92.0%		
a11-a12-a13 (5)	Geobacter sulfurreducens	91.5%		
a27-a28-a29 (8)	Desulfoacinum infernum	88.4%	Clone S-31 (d-Protéo) ²	89.1 %
a40-a41-a42 (6)	Desulfivibrio aminophilus	85.3%	Clone S-34 (d-Protéo) ²	86.2%
G+ bas GC%				
a23.1	Kurthia zopfii	97.3%		
a25.1	Kurthia zopfii	97.2%		
a18.1 (22)	Kurthia gipsonii	94.4%	G+ bas GC% non identifié RS19 (non publié)	94.8%

TABLEAU 7 (suite 3)
Diversité des séquences d'ADNr16S contenues dans la banque cosmétique

Actinomycètes				
a33.1	Cellulomonas sp	99.5%		
a14.7	Streptosporangium longisporum	99.8%		
a21.7	Arthrobacter polychromogenes	99.2%		
a8-a9-a10 (7)	Arthrobacter oxydans	98.3%	actinomycète non identifié RSW1 (non publié)	98.5%
a27-a28-a29 (3)	Arthrobacter oxydans	98.9%	actinomycète non identifié RSW1 (non publié)	99.3%
Acidobacterium ?				
a43-a44-a45 (4)	Holophaga foetida	87.3%	Clone 32-10 (Acidobacterium phylum) ⁶	95.0%
a27-a28-a29 (12)	Desulfuromonas acetexigens	88.8%	Clone Sva0515 (Acidobacterium phylum) ⁶	91.0%
a37-a38-a39 (12)	Desulfuromonas palmitatis	90.3%	Clone Sva0515 (Acidobacterium phylum) ⁶	91.5%
a37-a38-a39 (14)	Halothermothrix orenii	87.5%	Clone ii3-7 (Acidobacterium phylum) ⁶	93.3%
a8-a9-a10 (9)	Pelobacter carbinolicus	86.5%	Clone ii3-15 (Acidobacterium phylum) ⁶	92.6%
a34-a35-a36 (10)	Nitrococcus mobilis	90.6%	clone RB43 (Acidobacterium phylum) ⁶	93.7%
Non classifié				
a22.1(19)	Aerothermobacter marianas	89.1%	Eubactérie non identifiée (non publié)	93.4%

¹ GONZALEZ et al (1996) - ² Zhou et al. (1997) - ³ Pederson et al (1996) - ⁴ Godon et al (1997) - ⁵ Hugenholtz et al (1998)
⁶ Ludwig (1997)

TABLEAU 9 : Séquences

Désignation	SEQ ID N°
Sondes et amorces	
FGPS431	1
FGPS122	2
FGPS350	3
FGPS643 (T)	4
FGPS643 (C)	5
R499	6
R500	7
C501	8
FGPS516	9
FGPS517	10
FGPS518	11
FGPS612	12
FGPS669	13
FGPS618	14
FGPS614	15
FGPS615	16
FGPS616	17
FGPS621	18
FGPS617	19
FGPS680	20
FGPS619	21
63f	22
1387r	23
Oligo-1 (Exemple 10)	24
Oligo-2 (Exemple 10)	25
A1	26
A2	27
B1	28
B2	29
Acides nucléiques PKS-I	
Amb9	30
Amb12	31
Ery19	32
A9b12	33
A23G1 1-1	34
A26G1 1-2	35
A26G1-10	36

TABLEAU 9 (suite 1): Séquences

Désignation	SEQ ID N°
A35 E4-16	37
A49F1-32	38
A17d2-3	39
A53F11-13	40
A53F11-14	41
A22A 2-11	42
A36E8-1	43
A52E8-2	44
Séquences d'acides aminés PKS-I	
Amb9	45
Amb12	46
Ery19	47
A9b12	48
A23G1 1-1	49
A26G1 1-2	50
A26G1-10	51
A35 E4-16	52
A49F1-32	53
A17d2-3	54
A53F11-13	55
A53F11-14	56
A22A 2-11	57
A36E8-1	58
A52E8-2	59
Séquences ADNr 16S	
a24.1(2),	60
a4.a6.a7 (7)	61
a52.a53.a5(15)	62
a49.a50.a51(11)	63
a4.a6.a7(14)	64
a30.a31.a32(7)	65
a37.a38.a39(6)	66
a46.a47.a48(14)	67
a49.a50.a51(1)	68
a52.a53.a5(8)	69
a8.a9.a10(13)	70
a1.a2.a3(13)	71
a43.a44.a45(10)	72
a27.a28.a29(5)	73

TABLEAU 9 (suite 2): Séquences

Désignation	SEQ ID N°
a23.1	74
a25.1	75
a18.1(22)	76
a33.1	77
a14.7	78
a21.7	79
a8.a9.a10(7)	80
a8.a9.a10(18)	81
a27.a28.a29(3)	82
a34.a35.a36(5)	83
a22.1(19)	84
a11.a12.a13(5)	85
a19.a20.a26(9)	86
a40.a41.a42(6)	87
a27.a28.a29(8)	88
a27.a28.a29(12)	89
a37.a38.a39(12)	90
a46.a47.a48(6)	91
a11.a12.a13(11)	92
a15.a16.a17(12)	93
a15.a16.a17(5)	94
a19.a20.a26(13)	95
a37.a38.a39(14)	96
a8.a9.a10(9)	97
a19.a20.a26(5)	98
a43.a44.a45(4)	99
a1.a2.a3(4)	100
a4.a6.a7(23)	101
a49.a50.a51(22)	102
a8.a9.a10(2)	103
a34.a35.a36(3)	104
a34.a35.a36(10)	105
a40.a41.a42(13)	106

TABLEAU 9 (suite 3) :**Séquences**

Désignation	SEQ ID N°
Amorces	
cos 1 n (exemple 5)	107
cos 2 n (exemple 5)	108
EV ϕ c 31I (exemple 15)	109
Bll ϕ c 31F (exemple 15)	110
Amorce 1 (exemple 15)	111
Amorce 2 (exemple 15)	112
Acides nucléiques PKS-I	
Cosmide a2641 (vecteur + insert brin (-))	113
Cosmide a2641 (insert - brin (+))	114
orf1	115
orf2	116
orf3	117
orf4	118
orf5	119
orf6	120
Séquences acides aminés PKS-I	
ORF1	121
ORF2	122
ORF3	123
ORF4	124
ORF5	125
ORF6	126

REFERENCES

- **Amann, R. I., W. Ludwig, and K.-H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143-169.
- **Atschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J.** (1997) " Gapped BLAST and PSI-BLAST : a nex generation of protein databses search programs " *Nucleic Acid Researchs* Vol 25 : 3389-3404
- **Atschul SF et al.,** 1990, *J. Mol Biol*, **215** : 403-410.
- **Bakken, L. R.** 1985. Separation and purification of bacteria from soil. *Appl. Environ. Microbiol.* **49**:1482-1487.
- **Bibb MJ, Findlay PR, Johnson MW,** The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences., *Gene* 30: 1-3, 157-66, Oct, 1984.
- **Biesiekierska-Galguen M.** (1997) " Atténuation biologique de contaminants xénobiotiques dans le sol - modèle lindane " Diplôme de DEA National de Toxicologie, Université Claude Bernard Lyon I.
- **Blondelet-Rouault MH, Weiser J, Lebrihi A, Branny P, Pernodet JL.** Institut de Genetique et Microbiologie, URA CNRS 2225, Universite Paris XI, Orsay, France. *Gene* 1997 May 6;190(2):315-7
- **Borchert S et al.,** 1992, *Microbiology Letters*, **92** : 175-180
- **BLONDELET-ROUAULT,** 1997, *Gene*, 315-317.

- **Boccard, F., Smokvina, T. Pernodet, J.L. Friedmann, A. & Guerineau M. (1989)** . The integrated conjugative plasmid pSAM2 of *Streptomyces ambifaciens* is related to temperate bacteriophages. *Embo J* 8,973-80
- **Chatzinotas A., Sandaa R-A., Schönhuber W., Amanna R., Daae F.L., Torsvik V., Zeyer J., Hahn D. (1998)** " Analysis of broad-scale differences in microbial community composition of two pristine forest soils " *Systematic and Applied Microbiology* Vol 21 : 579-587
- **Clegg, C. D., K. Ritz, and B. S. Griffiths.** 1997. Direct extraction of microbial community DNA from humified upland soils. *Lett. Appl. Microbiol.* 25:30-33.
- **Clerc-Bardin, S., J.-L. Pernodet, A. Frostegård, and P. Simonet.** Development of a conditional suicide system for a *Streptomyces lividans* strain and its use to investigate conjugative transfer in soil. Submitted.
- **Elledge SJ, Mulligan JT, Ramer SW, Spottswood M, Davis RW.** Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030. *Proc Natl Acad Sci U S A* 1991 Mar 1;88(5):1731-5
- **Engelen, B., K. Meinken, F. Von Wintzingerode, H. Heuer, H.-P. Malkomes, and H. Backhaus.** 1998. Monitoring impact of a pesticide treatment on bacterial soil communities by metabolic and genetic fingerprinting in addition to conventional testing procedures. *Appl. Environ. Microbiol.* 64:2814-2821.
- **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and *rna* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* 61:2798-2801.
- **Faulkner D.V., Jurka J. (1988)** " Multiple Aligned Sequence Editor (MASE) " *Trends in Biochemical Sciences* Vol 13 : 321-322
- **FRENGEN et al., 1999, Genomics, 58: 250-258.**

- **Frostegård, A., Tunlid, A., and Bååth, E.** 1991. Microbial biomass measured as total lipid phosphate in soils of different organic content. *J. Microbiol. Meth.* **14**:151-163.
- **Giddings, G.** 1998. The release of genetically engineered micro-organisms and viruses into the environment. *New Phytol.* **140**:173-184.
- **Gladek, A., and J. Zakrzewska.** 1984. Genome size of *Streptomyces*. *FEMS Microbiol. Lett.* **24**:73-76.
- **Gribskov M, Devereux J, Burgess RR,** The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression., *Nucleic Acids Res* 12: 1 Pt 2, 539-49, Jan 11, 1984.
- **Guiney et al.,** 1983, *Proc. Natl. Acad. Sci USA*, **(12)** : 3595-3598.
- **Gourmelen, A. Blondelet-Rouault, M.H. & Pernodet, J.L.** (1998). Characterization of a glycosyl transferase inactivating macrolides, encoded by *gimA* from *Streptomyces ambofaciens*, *Antimicrob Agents Chemother* **42**, 2612-9.
- **Hayakawa, M., and H. Nonomura.** 1987. Humic acid-vitamin agar, a new medium for the selective isolation of soil actinomycetes. *J. Ferment. Technol.* **65**:501-509.
- **Hayakawa, M., Ishizawa K., and H. Nonomura.** 1988. Distribution of rare actinomycetes in Japanese soils. *J. Ferment. Technol.* **66**:367-373.
- **Hickey, R. J., and H. D. Tresner.** 1952. A cobalt containing medium for sporulation of *Streptomyces* species. *J. Bacteriol.* **64**:891-892.
- **Hintermann, G., R., Crameri, Kieser, T., and R. Hütter.** 1981. Restriction analysis of the *Streptomyces glaucescens* genome by agarose gel electrophoresis. *Arch. Microbiol.* **130**:218-222.

- **Holben, W. E., J. K. Jansson, B. K. Chelm, and J. M. Tiedje. 1988.** DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Appl. Environ. Microbiol.* **54**:703-711.
- **Hong Fu et al., 1995, Molecular diversity, 1 : 121-124**
- **Hopwood DA, Bibb M J, Chater K F, Kieser T., Bruton C.J., Kieser H.M., Lydiate D.J., Smith C.P., Ward J.M. and Scrempf H. 1985.** Genetic Manipulation of *Streptomyces*. A Laboratory manual. The John Innes Foundation, Norwich, U.K.
- **Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schrempf. 1985.** Genetic manipulation of streptomyces - a laboratory manual. The John Innes Foundation, Norwich, United Kingdom.
- **Hohm B. and Collins J., 1980, Gene, 11:291-298.**
- **Horinouchi S., Malpartida F., Hopwood D. et Beppu T., Mol. Gen. Genet. (1989) 215 :355-357.**
- **Imai R., Nagata Y., Fukuda M., Takagi M., Yano K. (1991) " Molecular cloning of a *Pseudomonas paucimobilis* gene encoding a 17-kilodalton ptypeptide that eliminates HCl molecules from γ -Hexachlorocyclohexane " *Journal of Bacteriology* Vol 17 ", No21 : 6811-6819**
- **Jacobsen, C. S., and O. F. Rasmussen. 1992.** Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Appl. Environ. Microbiol.* **58**:2458-2462.
- **Jae-Hyuk Y.U. and Leonard T.J.,1995.** Sterigmetscytin biosynthesis in *Aspergillus nidulans* requires a ... type I polyketide synthase. *J. Bacteriol*, (August) : 4792-4800.

- **Ka, J. O., W. E. Holben, and J. M. Tiedje.** 1994. Analysis of competition in soil among 2,4-dichlorophenoxyacetic acid-degrading bacteria. *Appl. Environ. Microbiol.* **60**:1121-1128.
- **Kah-Tong S et al.,** 1997, *J Bacteriol*, G179(23) : **7360-7368**
- **Kimura M.** (1980) " A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences " *Journal of Molecular Evolution* Vol 16 : 111-120
- **Kuske, C. R., K. L. Banton, D. L. Adorada, P. C. Stark, K. K. Hill, and P. J. Jackson.** 1998. Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl. Environ. Microbiol.* **64**:2463-2472.
- **Lacalle RA, Pulido D, Vara J, Zalacain M, Jimenez A.** Centro de Biologia Molecular (CSIC-UAM), Universidad Autonoma, Canto Blanco, Madrid, Spain. *Gene* 1989 Jul 15;79(2):375-80
- **Lee, S.-Y., J. Bollinger, D. Bezdicek, and A. Ogram.** 1996. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. *Appl. Environ. Microbiol.* **62**:3787-3793.
- **Leff, L. G., J. R. Dana, J. V. McArthur, and L. J. Shimkets.** 1995. Comparison of methods of DNA extraction from stream sediments. *Appl. Environ. Microbiol.* **61**:1141-1143.
- **Liesack, W., and E. Stackebrandt.** 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* **174**:5072-5078.
- **Liesack, W., P. H. Janssen, F. A. Rainey, N. L. Ward-Rainey, and E. Stackebrandt.** 1997. Microbial diversity in soil: the need for a combined approach using molecular and cultivation techniques. *In* J. D. Van Elsas, J.

T. Trevors, and E. M. H. Wellington (ed.), Modern soil microbiology, Marcel Dekker, Inc., New York. (p 375-439)

- **Lorentz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Reviews* 58:563-602.
- **Maidak B.L., Cole J.R., Parker C.T., Garrity G.M., Larsen N., Li B., Lilburn T.G., McCaughey M.J., Olsen G.J., Overbeek R., Pramanik S., Schmidt T.M., Tiedje J.M., Woese C.R.** (1999) " A new project of the RDP (Ribosomal Database Project) " *Nucleic Acids Research* Vol 27 : 171-173
- **Mazodier P. et al.,** 1989, *J. Bacteriol.*, **171(6)** : 3583-3585.
- **Moré, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse, and E. L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl. Environ. Microbiol.* **60**:1572-1580.
- **Murakami T, Holt TG, Thompson CJ,** Unité de Génie Microbiologique, Institut Pasteur, Paris, France. *J. Bacteriol* 1989 Mar;171(3):1459-66
- **Nagata Y., Hatta T., Imai R., Kimbara K., Fukuda M., Yano K., Takagi M.** (1993) " Purification and characterization of γ -Hexachlorocyclohexane (γ -HCH)dehydrochlorinase (LinA) from *Pseudomonas paucimobilis* " *Bioscience, Biotechnology and Biochemistry* Vol 57 No 9 : 1582-1583
- **Nalin R., Simonet P., Vogel T.M., Normand P.** (1999) " *Rhodanobacter lindaniclasticus* gen.nov., sp., nov., a lindane-degrading bacterium " *International Journal of Systematic Bacteriology* Vol 49 : 19-23
- **Nesme, X., C. Picard, and P. Simonet.** 1995. Specific DNA sequences for detection of soil bacteria. *In* J. T. Trevors, and J. D. van Elsas (ed.), *Nucleic acids in the environment, methods and application.* Springer Lab Manual. (p 111-139)

- Nilsson B, Uhlen M, Josephson S, Gatenbeck S, Philipson L. *Nucleic Acids Res* 1983 Nov 25;11(22):8019-30
- Normand P. et al., 1995, *Océanis*, 21(1) : 31-56
- Ogram, A. V., M. L. Mathot, J. B. Harsh., J. Boyle, and C. A. Pettigrew, JR. 1994. Effects of DNA polymer length on its adsorption to soils. *Appl. Environ. Microbiol.* 60:393-396.
- Ogram, A., G. S. Sayler, and T. Barkay. 1987. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* 7:57-66.
- Olsen, R. A., and Bakken, L. R. 1987. Viability of soil bacteria: optimization of the plate-counting technique. *Microb. Ecol.* 13:59-74.
- Paget, E., L. Jocteur Monrozier, and P. Simonet. 1992. Adsorption of DNA on clay minerals: protection against DNaseI and influence on gene transfer. *FEMS Microbiol. Lett.* 97:31-40.
- Patra, G., P. Sylvestre, V. Ramisse, J. Thérasse, and J.-L. Guesdon. 1996. Isolation of a specific chromosomal DNA sequence of *Bacillus anthracis* and its possible use in diagnosis. *FEMS Immunol. Medical Microbiology* 15:223-231.
- Pernodet J.L. Fish, S. Blondelet-Rouault, M.H. & Cundliffe, E. (1996). The macrolide-lincosamide-streptogramin B resistance phenotypes characterized by using a specifically deleted, antibiotic-sensitive strain of *Streptomyces lividans*. *Antimicrob Agents Chemother* 40, 581, 5.
- Pernodet J.L. , Gourmelen, A., Blondelet-Rouault, M.H. & Cundliffe, E. (1999). Dispensable ribosomal resistance to spiramycin conferred by *smA* in the spiramycin producer *Streptomyces ambofaciens*. 145, 2355-64.
- Picard, C., C. Ponsonnet, X. Nesme, and P. Simonet. 1992. Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Appl. Environ. Microbiol.* 58:2717-2722.

- **Preud'homme, J., Belloc, A., Charpentié, Y., and Tarridec, P.** 1965. Un antibiotique formé de deux groupes de composants à synergie d'action : la pristnamycine C. R. Acad. Sci. 260 :1309-1312.
- **Priemé, A., J. I. B. Sitaula, A. K. Klemetsson, and L. R. Bakken.** 1996. Extraction of methane-oxidizing bacteria from soil particles. FEMS Microbiol. Ecol. **21**: 59-68.
- **Prosser, J.** 1994. Molecular marker systems for detection of genetically engineered micro-organisms in the environment. Microbiol. **140**:5-17.
- **Raynal A, Tuphile K, Gerbaud C, Luther T, Guérineau M, Pernodet JL ;** Laboratoire de Biologie et Génétique Moléculaire, Institut de Génétique et Microbiologie, URA CNRS 2225, Université Paris-Sud, Orsay, France. Mol Microbiol 1998 Apr;**28**(2):333-42
- **Raynald A. Tuphile, K. Gerbaud, C., Luther, T. Guérineau, M. & PERNODET, J.L. (1998).** Structure of the chromosomal insertion site for pSAM2: functional analysis in Escherichia coli. Mol. Microbiol **28**, 333-42.
- **Richard, G. M.** 1974. Modifications of the diphenylamine reaction giving increased sensitivity and simplicity in the estimation of DNA. Analytical Biochem. **57**:369-376.
- **Romanowski, G., M. G. Lorentz, and W. Wackernagel.** 1993. Use of polymerase chain reaction and electroporation of Escherichia coli to monitor the persistence of extracellular plasmid DNA introduced into natural soils. Appl. Environ. Microbiol. **59**:3438-3446.
- **Saitou N., Nei M. (1987)** " The Neighbour-Joining method : a new method for reconstructing phylogentic trees " *Molecular and Biological Evolution* Vol 2 : 112-118

- **Sambrook J., Fritsch E. F. et Maniatis T. 1996.** Molecular cloning : a laboratory manual, 2nd ed. Cold spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- **Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989.** Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
- **Senoo K., Wada H. (1989)** " Isolation and identification of an aerobic ?-HCH-decomposing bacterium from soil " *Soil Science, Plant Nutrition* Vol 35, No 1 : 79-87.
- **Sezonov, G., Blanc, V., Bamas-Jacques, N., Friedmann, A. Pernodet, J.L. & Guerineau, M.(1997).** Complete conversion of antibiotic precursor to pristinamycin IIA by overexpression of *Streptomyces pristinae* biosynthetic genes. *Nat Biotechnol* 15,349-53.
- **Shirling, E. B., and D. Gottlieb. 1966.** Methods for characterization of *Streptomyces* species. *Int. J. Syst. Bacteriol.* 16:313-340.
- Shizuga et al., 1992, *Proc. Natl. Acad. Sci USA*, 89 : 8794-8797.
- **Siefert, J. L., and G. E. Fox. 1998.** Phylogenetic mapping of bacterial morphology. *Microbiology* 144:2803-2808.
- **Simonet, P., P. Normand, A. Moiroud, and R. Bardin. 1990.** Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. *Arch. Microbiol.* 153:235-240.
- **Smalla, K., N. Cresswell, L. Mendonca-Hagler, A. Wolters, and D. J. van Elsas. 1993.** Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. *J. Appl. Bacteriol.* 74:78-85.
- **SOSIO M. et al. 2000, Nature Biotechnology, vol.18,:343-345.**

- **Smit, E., P. Leeflang, and K. Wernars.** 1997. Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. *FEMS Microbiol. Ecol.* **23**:249-261.
- **Smokvina T, Mazodier P, Boccard F, Thompson CJ, Guerinéau M.** Laboratoire de Biologie et Genetique Moleculaire, Université Paris-Sud, Orsay, France. *Gene* 1990 Sep 28;94(1):53-9
- **Smolovina, T., Mazodier, P. Boccard, F. Thompson, C.J. & Guerinéau, M. (1990).** Construction of a series of pSAM2-based integrative vectors for use in actinomycetes. *Gene* 94, 53-9.
- **Stackebrandt, E.** 1988. Phylogenetic relationships vs. phenotypic diversity: how to achieve a phylogenetic classification system of the eubacteria. *Can. J. Microbiol.* **34**:552-556.
- **Staneck, J. L., and G. D. Roberts.** 1974. Simplified approach to identification of aerobic Actinomycetes by thin-layer chromatography. *Appl. Microbiol.* **28**:226-231.
- **Stapleton, R. D., S. Ripp, L. Jimenez, S. Cheol-Koh, J. T. Fleming, I. R. Gregory, and G. S. Sayler.** 1998. Nucleic acid analytical approaches in bioremediation: site assessment and characterization. *J. Microbiol. Methods* **32**:165-178.
- **Steffan, R. J., J. Goksøyr, A. K. Bej, and R. Atlas.** 1988. Recovery of DNA from soils and sediments. *Appl. Environ. Microbiol.* **54**:2908-2915.
- **Tebbe, C. C., and W. Vahjen.** 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl. Environ. Microbiol.* **59**:2657-2665.

- **Tercero JA, Espinosa JC, Lacalle RA, Jimenez A .** Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas, Madrid, Spain. *J Biol Chem* 1996 Jan 19;271(3):1579-90
- **Thomas J-C., Berger F., Jacquier M., Bernillon D., Baud-Grasset F., Truffaut N., Normand P., Vogel T.M., Simonet P.** (1996) " Isolation and Characterisation of a novel ?-Hexachlorocyclohexane-degrading bacterium " *Journal of Bacteriology* Vol 178, No20 : 6049-6055
- **Torsvik, V. L.** 1980. Isolation of bacterial DNA from soil. *Soil Biol. Biochem.* 12:15-21.
- **Torsvik, V., R. Sørheim, and J. Goksøyr.** 1996. Total bacterial diversity in soil and sediment communities – a review. *J. Ind. Microbiol.* 17:170-178.
- **Tsai, Y.-L., and B. Olson.** 1991. Rapid method for direct extraction of DNA from soil and sediments. *Appl. Environ. Microbiol.* 57:1070-1074.
- **Umeyama T., Tanabe Y., Aigle B.D. et Horinuochi S.,** FEMS (1996) 144 :177-184.
- **Van Elsas, J. D., G. F. Duarte, A. S. Rosado, and K. Smalla.** 1998. Microbiological and molecular biological methods for monitoring microbial inoculants and their effects in the soil environment. *J. Microbiol. Methods* 32:133-154.
- **Van Elsas, J. D., V. Mäntynen, and A. C. Wolters.** 1997. Soil DNA extraction and assessment of the fate of *Mycobacterium chlorophenolicum* strain PCP-1 in different soils by 16S ribosomal RNA gene sequence based most-probable-number PCR and immunofluorescence. *Biol. Fert. Soils* 24:188-195.
- **Volff JN et al.,** 1996, *Mol. Microbiol.*, 21(5) : 1037-1047.
- **Volossiuk, T., E. J. Robb, and R. N. Nazar.** 1995. Direct DNA extraction for PCR-mediated assays. *Appl. Environ. Microbiol.* 61:3972-3976.

- **Wahl GM, Lewis KA, Ruiz JC, Rothenberg B, Zhao J, Evans GA.** Proc Natl Acad Sci U S A 1987 Apr;84(8):2160-4
- **Waksman, S. A.** 1961. Williams and Wilkins (ed.) The actinomycetes. Classification, identification and description of genera and species.Vol 2. Baltimore.
- **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 344:63-65.
- **Widmer, F., R. J. Seidler, and L. S. Watrud.** 1996. Sensitive detection of transgenic plant marker gene persistence in soil microcosms. Mol. Ecol. 5:603-613.
- **Williams, S.T., R. Locci, A. Beswick, D. I. Kurtböke, V. D. Kuznetsov, F. J. Le Monnier, P. F. Long, K. A. Maycroft, R. A. Palma, B. Petrolini, S. Quaroni, J. I. Todd, and M. West.** 1993. Detection and identification of novel actinomycetes. Res. Microbiol. 144:653-656.
- **Wilson, I. G.** 1997. Inhibition and facilitation of nucleic acid amplification. Appl. Environ. Microbiol. 63:3741-3751.
- **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. 51:221-271.
- Yannish-Perron et al., 1985 , Gene, 33(1) : 103-119.
- **Zaslavsky, B. Y.** 1995. Separation of biomolecules, p. 503-667. *In* Aqueous two-phase partitioning. Boris Y. Zaslavsky (ed.) Physical Chemistry and Bioanalytical Applications, Marcel Dekker, Inc., New York.
- **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62:316-322.

REVENDEICATIONS

1. Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de sol contenant des organismes, ledit procédé comprenant la succession d'étapes suivante :

- I (a) Obtention de micro-particules par broyage d'un échantillon de sol préalablement séché ou dessiqué , puis mise en suspension des micro-particules dans un milieu tampon liquide ; et
 (b) extraction des acides nucléiques présents dans les micro-particules ; et
 (c)- passage de la solution contenant les acides nucléiques sur un tamis moléculaire, puis récupération des fractions d'élution enrichies en acides nucléiques et passage des fractions d'élution enrichies en acides nucléiques sur un support de chromatographie d'échange d'anions, puis récupération des fractions d'élution contenant les acides nucléiques purifiés.

2. Procédé de préparation d'une collection d'acides nucléiques à partir d'un échantillon de l'environnement contenant des organismes, ledit procédé comprenant la succession d'étapes suivante :

- II (i) Obtention d'une suspension par dispersion de l'échantillon de l'environnement en milieu liquide puis homogénéisation de la suspension par agitation douce; et
 (ii) séparation des organismes et des autres constituants minéraux et/ou organiques de la suspension homogène obtenue à l'étape (i) par centrifugation sur un gradient de densité ; et
 (iii) lyse des organismes séparés à l'étape (ii) et extraction des acides nucléiques; et
 (iv) purification des acides nucléiques sur un gradient de chlorure de césium.

3. Procédé selon la revendication 1, caractérisé en ce que l'étape I- (a) est suivie d'une étape complémentaire de :

- traitement des micro-particules en suspension dans un milieu tampon liquide par sonication ;

4. Procédé selon la revendication 1, caractérisé en ce que l'étape I-(a) est suivie des étapes complémentaires suivantes :

- traitement des micro-particules en suspension dans un milieu tampon liquide par sonication ;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- addition de SDS
- récupération des acides nucléiques.

5. Procédé selon la revendication 1, caractérisé en ce que l'étape I-(a) est suivie des étapes complémentaires suivantes :

- homogénéisation des micro-particules à l'aide d'une étape de mixage violent (vortex) suivie d'une étape de simple agitation ;
- congélation de la suspension homogène suivie d'une décongélation ;
- traitement par sonication de la suspension après décongélation ;
- incubation de la suspension à 37°C après sonication en présence de lysozyme et d'achromopeptidase;
- addition de SDS;

6. Procédé selon l'une des revendications 1 à 5 caractérisé en ce que les acides nucléiques sont des molécules d'ADN.

7. Procédé de préparation d'une collection de vecteurs recombinants, caractérisé en ce que les acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6 sont insérés dans un vecteur de clonage et/ou d'expression.

8. Procédé selon la revendication 7, caractérisé en ce que les acides nucléiques sont séparés en fonction de leur taille préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

9. Procédé selon la revendication 7, caractérisé en ce que la taille moyenne des acides nucléiques est rendue sensiblement uniforme par

rupture physique, préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

10. Procédé selon la revendication 7, caractérisé en ce que le vecteur de clonage et/ou d'expression est du type plasmide.

11. Procédé selon la revendication 7, caractérisé en ce que le vecteur de clonage et/ou d'expression est du type cosmide.

12. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide répliatif chez *E. coli* et intégratif chez *Streptomyces*.

13. Procédé selon la revendication 12, caractérisé en ce qu'il s'agit du cosmide pOS700I.

14. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide conjugatif et intégratif chez *Streptomyces*.

15. Procédé selon la revendication 14, caractérisé en ce que le cosmide est choisi parmi les cosmides pOSV303, pOSV306 et pOSV307.

16. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide répliatif à la fois chez *E. coli* et chez *Streptomyces*.

17. Procédé selon la revendication 16, caractérisé en ce qu'il s'agit du cosmide pOS 700R.

18. Procédé selon la revendication 11, caractérisé en ce qu'il s'agit d'un cosmide répliatif chez *E. coli* et *Streptomyces* et conjugatif chez *Streptomyces*.

19. Procédé selon la revendication 7, caractérisé en ce que le vecteur de clonage et/ou d'expression est du type BAC.

20. Procédé selon la revendication 19, caractérisé en ce qu'il s'agit d'un vecteur BAC intégratif et conjugatif chez *Streptomyces*.

21. Procédé selon la revendication 20, caractérisé en ce que le vecteur est choisi parmi les vecteurs BAC pOSV403, pMBD-1, pMBD-2, pMBD-3, pMBD-4, pMBD-5 et pMBD-6.

22. Procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression, caractérisé en ce que l'étape d'insertion d'un acide nucléique dans le vecteur de clonage et/ou d'expression comprend les étapes suivantes :

- ouvrir le vecteur de clonage et/ou d'expression à un site de clonage choisi, à l'aide d'une endonucléase de restriction appropriée ;
- ajouter un premier acide nucléique homopolymérique à l'extrémité 3' libre du vecteur ouvert ;
- ajouter un second acide nucléique homopolymérique, de séquence complémentaire au premier acide nucléique homopolymérique, à l'extrémité 3' libre de l'acide nucléique de la collection à insérer dans le vecteur;
- assembler l'acide nucléique du vecteur et l'acide nucléique de la collection par hybridation du premier et du second acide nucléique homopolymérique de séquences complémentaires l'une de l'autre;
- refermer le vecteur par ligation.

23. Procédé selon la revendication 22, caractérisé en ce que :

- le premier acide nucléique homopolymérique est de séquence poly(A) ou poly (T) ; et
- le second acide nucléique homopolymérique est de séquence poly(T) ou poly(A).

24. Procédé de préparation d'un vecteur recombinant selon l'une des revendications 22 ou 23, caractérisé en ce que la taille de l'acide nucléique à insérer est d'au moins 100 kilobases, préférentiellement d'au moins 200 kilobases.

25. Procédé de préparation d'un vecteur recombinant selon l'une des revendications 22 à 24, caractérisé en ce que l'acide nucléique à insérer est contenu dans la collection d'acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6.

26. Procédé de préparation d'un vecteur recombinant de clonage et/ou d'expression, caractérisé en ce que l'étape d'insertion d'un d'acide nucléique dans le vecteur de clonage et/ou d'expression comprend les étapes suivantes :

- création de bouts francs sur les extrémités de l'acide nucléique de la collection par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes ;
- ouverture du vecteur de clonage et/ou d'expression à un site de clonage choisi, à l'aide d'une endonucléase de restriction appropriée ;
- création de bouts francs aux extrémités de l'acide nucléique du vecteur par élimination des séquences 3' sortantes et remplissage des séquences 5' sortantes, puis déphosphorylation des extrémités 5' ;
- Addition d'adaptateurs oligonucléotidiques complémentaires ;
- insertion de l'acide nucléique de la collection dans le vecteur par ligation.

27. Procédé de préparation d'un vecteur recombinant selon la revendication 26, caractérisé en ce que la taille de l'acide nucléique à insérer est d'au moins 100 kilobases, préférentiellement d'au moins 200 kilobases.

28. Procédé de préparation d'un vecteur recombinant selon l'une des revendications 26 ou 27, caractérisé en ce que l'acide nucléique à insérer est contenu dans la collection d'acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6.

29. Procédé selon l'une des revendications 22 à 28, caractérisé en ce que les acides nucléiques sont insérés tels quels, sans traitement par une ou plusieurs endonucléases de restriction préalablement à leur insertion dans le vecteur de clonage et/ou d'expression.

30. Collection d'acides nucléiques constituée des acides nucléiques obtenus par le procédé selon l'une des revendications 1 à 6.

31. Acide nucléique caractérisé en ce qu'il est contenu dans la collection d'acides nucléiques selon la revendication 30.

32. Acide nucléique selon la revendication 31, caractérisé en ce qu'il comprend une séquence nucléotidique codant au moins un opéron, ou une partie d'un opéron.

33. Acide nucléique selon la revendication 32, caractérisé en ce que l'opéron code pour la totalité ou une partie d'une voie métabolique.

34. Acide nucléique selon la revendication 33, caractérisé en ce que la voie métabolique est la voie de synthèse des polykétides.

35. Acide nucléique selon la revendication 34, caractérisé en ce qu'il est choisi parmi les polynucléotides comprenant les séquences SEQ ID N° 30 à 44 et SEQ ID N° 115 à 120.

36. Acide nucléique selon la revendication 31, caractérisé en ce qu'il comprend la totalité d'une séquence nucléotidique codant pour un polypeptide

37. Acide nucléique selon l'une des revendications 31 à 36, caractérisé en ce qu'il est d'origine procaryote.

38. Acide nucléique selon la revendication 37, caractérisé en ce qu'il provient d'une bactérie ou d'un virus.

39. Acide nucléique selon l'une des revendications 31 à 33 et 36, caractérisé en ce qu'il est d'origine eucaryote.

40. Acide nucléique selon la revendication 39, caractérisé en ce qu'il provient d'un champignon, d'une levure, d'une plante ou d'un animal.

41. Vecteur recombinant caractérisé en ce qu'il est choisi parmi les vecteurs recombinants suivants :

- a) un vecteur comprenant un acide nucléique selon l'une des revendications 35 à 40;
- b) un vecteur obtenu selon le procédé de l'une des revendications 22 à 25 et 29;
- c) un vecteur obtenu selon le procédé de l'une des revendications 26 à 29.

42 Vecteur caractérisé en ce qu'il s'agit du cosmide pOS 700I.

43. Vecteur caractérisé en ce qu'il s'agit du cosmide pOSV303.

44. Vecteur caractérisé en ce qu'il s'agit du cosmide pOSV306.

45. Vecteur caractérisé en ce qu'il s'agit du cosmide pOSV307.

46. Vecteur caractérisé en ce qu'il s'agit du cosmide pOS 700R.

47. Vecteur caractérisé en ce qu'il s'agit du vecteur BAC pOSV403.

48. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-1.

49. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-2

50. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-3.

51. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-4.

52. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-5.

53. Vecteur caractérisé en ce qu'il s'agit du vecteur pMBD-6.

54. Collection de vecteurs recombinants tels qu'obtenus selon le procédé de l'une des revendications 7 à 21, 25 et 28.

55. Vecteur recombinant de clonage et/ou d'expression caractérisé en ce qu'il est contenu dans la collection de vecteurs recombinants selon la revendication 54.

56. Cellule hôte recombinante comprenant un acide nucléique selon l'une des revendications 31 à 40 ou un vecteur recombinant selon la revendication 55.

57. Cellule hôte recombinante selon la revendication 56, caractérisée en ce qu'il s'agit d'une cellule procaryote ou eucaryote.

58. Cellule hôte recombinante selon la revendication 57, caractérisée en ce qu'il s'agit d'une bactérie.

59. Cellule hôte recombinante selon la revendication 58, caractérisée en ce qu'il s'agit d'une bactérie choisie parmi *E. coli* et *Streptomyces*.

60. Cellule hôte recombinante selon la revendication 58, caractérisée en ce qu'il s'agit d'une levure ou d'un champignon filamenteux.

61. Collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutives de la collection comprenant un acide nucléique de la collection d'acides nucléiques selon la revendication 30.

62. Collection de cellules hôtes recombinantes, chacune des cellules hôtes constitutives de la collection comprenant un vecteur recombinant selon l'une des revendications 41 ou 55.

63. Procédé de détection d'un acide nucléique de séquence nucléotidique déterminée, ou de séquence nucléotidique structurellement apparentée à une séquence nucléotidique déterminée, dans une collection

de cellules hôtes recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- mettre en contact la collection de cellules hôtes recombinantes avec un couple d'amorces hybridant avec la séquence nucléotidique déterminée ou hybridant avec la séquence nucléotidique structurellement apparentée à une séquence nucléotidique déterminée ;
- réaliser au moins trois cycles d'amplification ;
- détecter l'acide nucléique éventuellement amplifié..

64. Procédé de détection d'un acide nucléique de séquence nucléotidique déterminée, ou de séquence nucléotidique structurellement apparentée à une séquence nucléotidique déterminée, dans une collection de cellules hôtes recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- mettre en contact la collection de cellules hôtes recombinantes avec une sonde hybridant avec la séquence nucléotidique déterminée ou hybridant avec une séquence nucléotidique structurellement apparentée à la séquence nucléotidique déterminée ;
- détecter l'hybride éventuellement formé entre la sonde et les acides nucléiques compris dans les vecteurs de la collection.

65. Procédé pour identifier la production d'un composé d'intérêt par une ou plusieurs cellules hôtes recombinantes dans une collection de cellules hôtes recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié ;
- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules hôtes recombinantes cultivées.

66 Procédé pour sélectionner une cellule hôte recombinante produisant un composé d'intérêt dans une collection de cellules hôtes

recombinantes selon l'une des revendications 61 ou 62, caractérisé en ce qu'il comprend les étapes suivantes :

- culture des cellules hôtes recombinantes de la collection dans un milieu de culture approprié ;
- détection du composé d'intérêt dans le surnageant de culture ou dans le lysat cellulaire d'une ou plusieurs des cellules hôtes recombinantes cultivées.
- sélection des cellules hôtes recombinantes produisant le composé d'intérêt.

67. Procédé pour la production d'un composé d'intérêt caractérisé en ce qu'il comprend les étapes suivantes :

- cultiver d'une cellule hôte recombinante sélectionnée selon le procédé de la revendication 66;
- récupérer et, le cas échéant, purifier, le composé produit par ladite cellule hôte recombinante.

68. Composé d'intérêt caractérisé en ce qu'il est obtenu selon le procédé de la revendication 67.

69. Composé selon la revendication 68, caractérisé en ce qu'il s'agit d'un polykétide.

70. Polykétide caractérisé en ce qu'il est produit grâce à l'expression d'au moins une séquence nucléotidique comprenant une séquence choisie parmi les séquences SEQ ID N°30 à 44 et SEQ ID N°115 à 120.

71. Composition comprenant un polykétide selon la revendication 69 ou 70.

72. Composition pharmaceutique comprenant une quantité pharmacologiquement active d'un polykétide selon la revendication 69 ou 70, en association avec un véhicule pharmaceutiquement compatible.

73. Procédé de détermination de la diversité des acides nucléiques contenus dans une collection d'acides nucléiques et tout particulièrement d'une collection d'acides nucléiques provenant d'un échantillon de l'environnement, préférentiellement d'un échantillon du sol, ledit procédé comprenant les étapes suivantes:

- mise en contact des acides nucléiques de la collection d'acides nucléiques à tester avec un couple d'amorces oligonucléotidiques hybridant à toute séquence d'ADN ribosomal 16 S bactérien;
- réalisation d'au moins trois cycles d'amplification;
- détection des acides nucléiques amplifiés à l'aide d'une sonde oligonucléotidique ou d'une pluralité de sondes oligonucléotidiques, chaque sonde hybridant spécifiquement avec une séquence d'ADN ribosomal 16 S commune à un règne, un ordre, une sous-classe ou un genre bactérien;
- le cas échéant, comparer les résultats de l'étape de détection précédente avec les résultats de détection, à l'aide de la sonde ou de la pluralité de sondes, d'acides nucléiques de séquence connue constituant une gamme étalon.

74. Procédé selon la revendication 73, caractérisé en ce que le couple d'amorces hybridant à toute séquence d'ADN ribosomal 16 S bactérien est constitué de l'amorce FGPS 612 (SEQ ID N°12) et de l'amorce FGPS 669 (SEQ ID N°13).

75. Procédé selon la revendication 73, caractérisé en ce que le couple d'amorces hybridant à toute séquence d'ADN ribosomal 16 S bactérien est constitué de l'amorce 63 f (SEQ ID N°22) et de l'amorce 1387 (SEQ ID N°23).

76. Acide nucléique comprenant une séquence nucléotidique d'ADNr 16S choisie parmi les séquences possédant au moins 99% d'identité en nucléotides avec les séquences SEQ ID N° 60 à SEQ ID N° 106.

77. Procédé de production d'une polykétide synthase de type I, ledit procédé de production comprenant les étapes suivantes:

- obtention d'une cellule hôte recombinante comprenant un acide nucléique codant pour une polykétide synthase de type I comprenant une séquence nucléotidique choisie parmi les séquences SEQ ID N°33 à SEQ ID N°44, SEQ ID N°30 à SEQ ID N°32 et SEQ ID N° 115 à SEQ ID N°120.

- culture des cellules hôtes recombinantes dans un milieu de culture approprié;

- récupération et, le cas échéant, purification de la polykétide synthase de type I à partir du surnageant de culture ou du lysat cellulaire.

78. Polykétide synthase comprenant une séquence d'acides aminés choisie parmi les séquences SEQ ID N°45 à 59 et SEQ ID N° 121 à SEQ ID N°126.

79. Anticorps dirigé contre une polykétide synthase selon la revendication 78.

80. Procédé de détection d'une polykétide synthase de type I ou d'un fragment peptidique de cette enzyme, dans un échantillon, ledit procédé comprenant les étapes de:

- a) mettre en contact un anticorps selon la revendication 79 avec l'échantillon à tester;

- b) détecter le complexe antigène/anticorps éventuellement formé.

81. Nécessaire de détection d'une polykétide synthase de type I dans un échantillon comprenant:

- a) un anticorps selon la revendication 79;

- b) le cas échéant, des réactifs nécessaires à la détection du complexe antigène/anticorps éventuellement formé.

Sol sec et tamisé

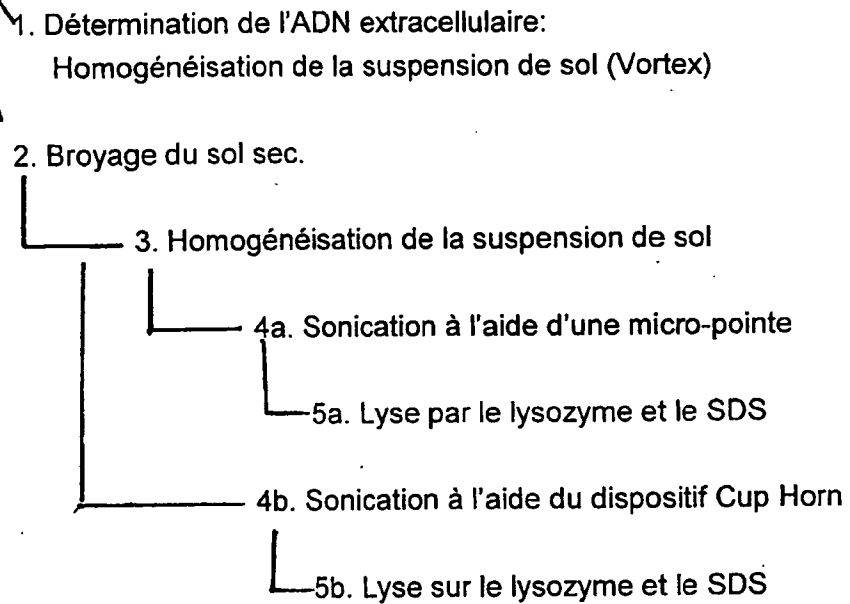


FIG. 1

2/38

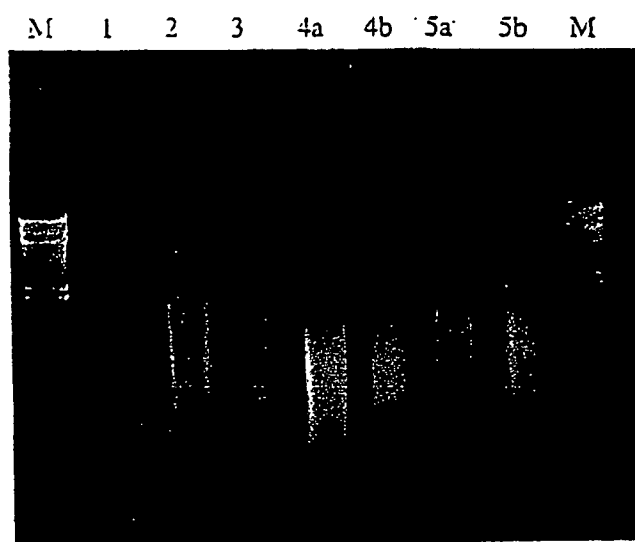


Figure 2

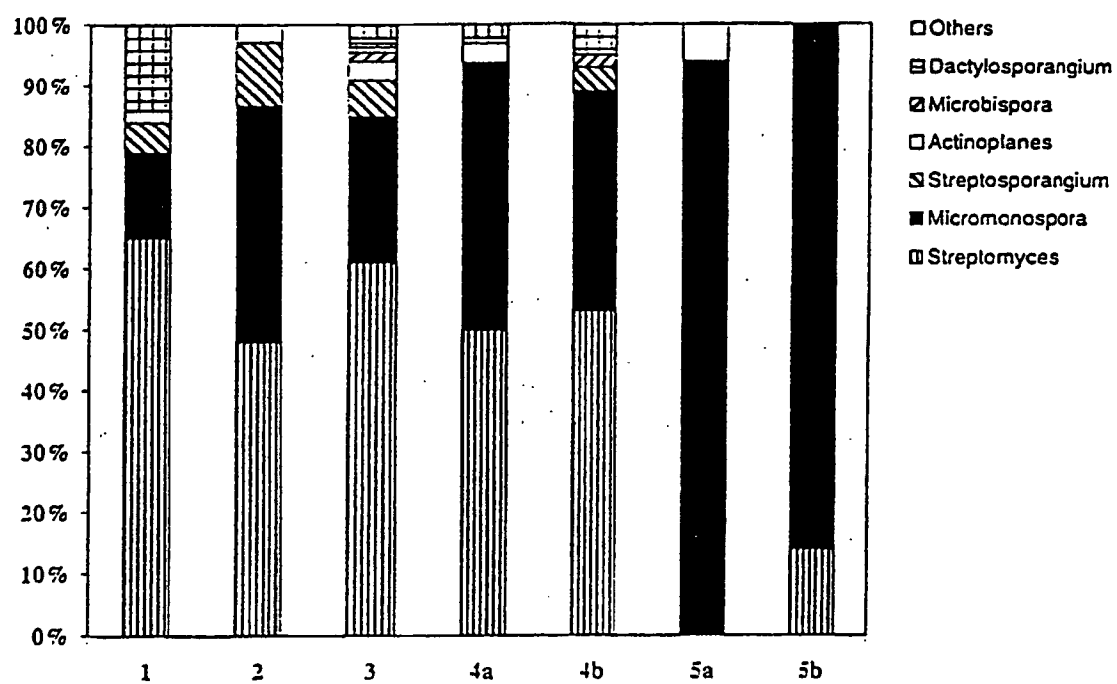


Figure 3

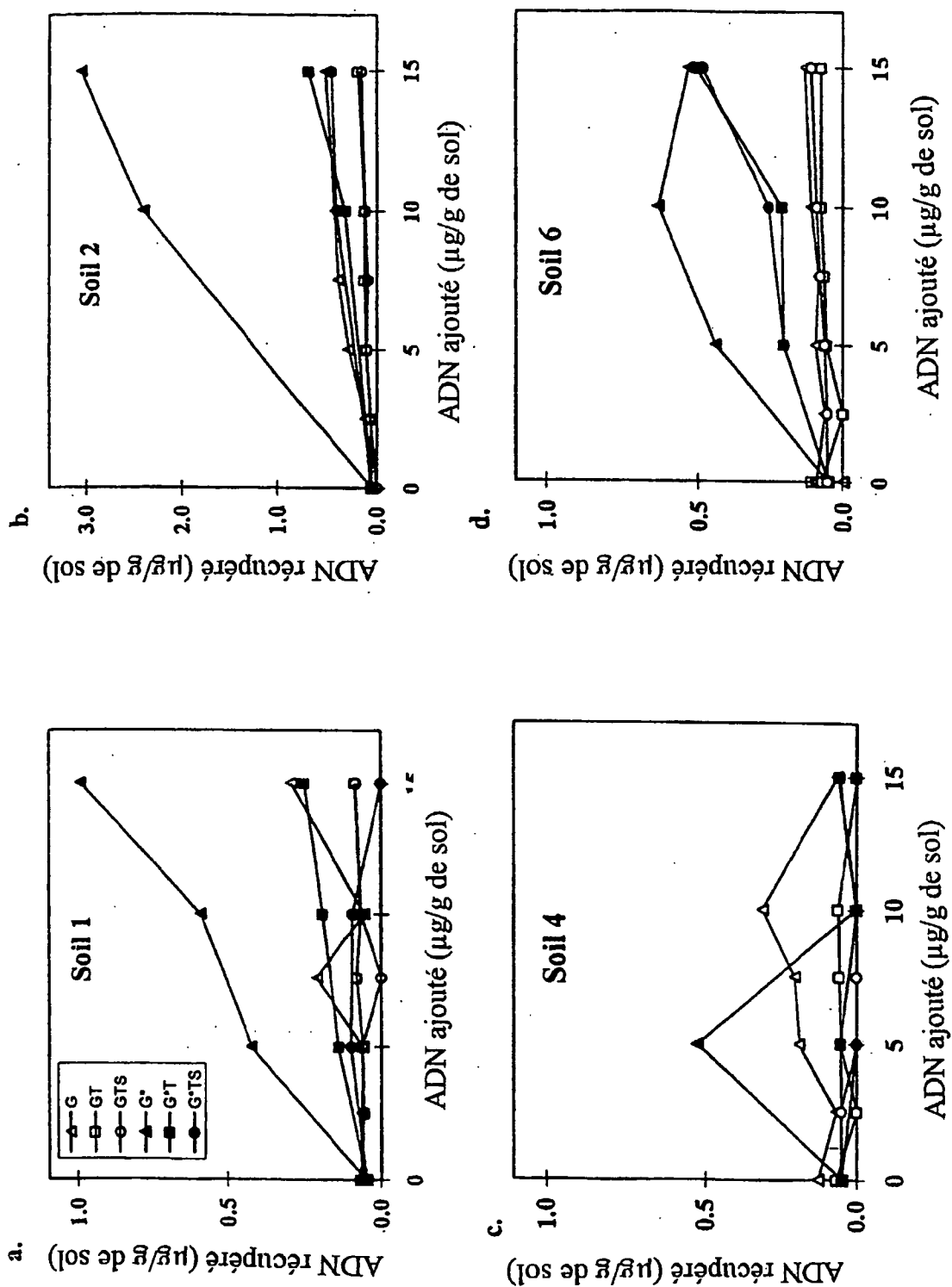


Figure 4

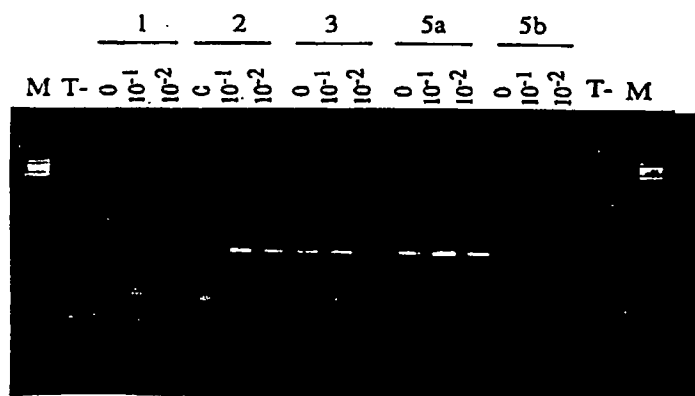


Figure 5

6/38

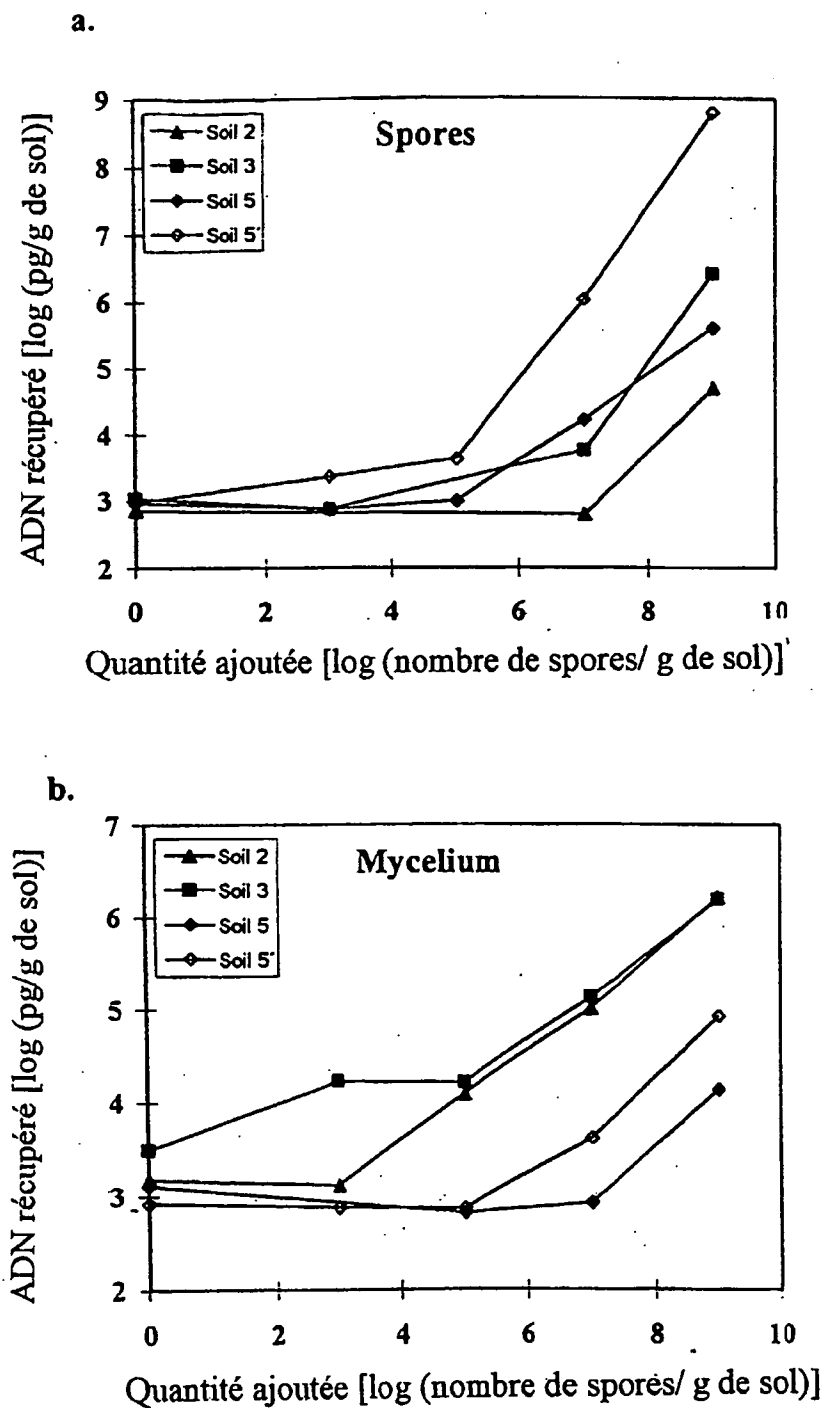


Figure 6

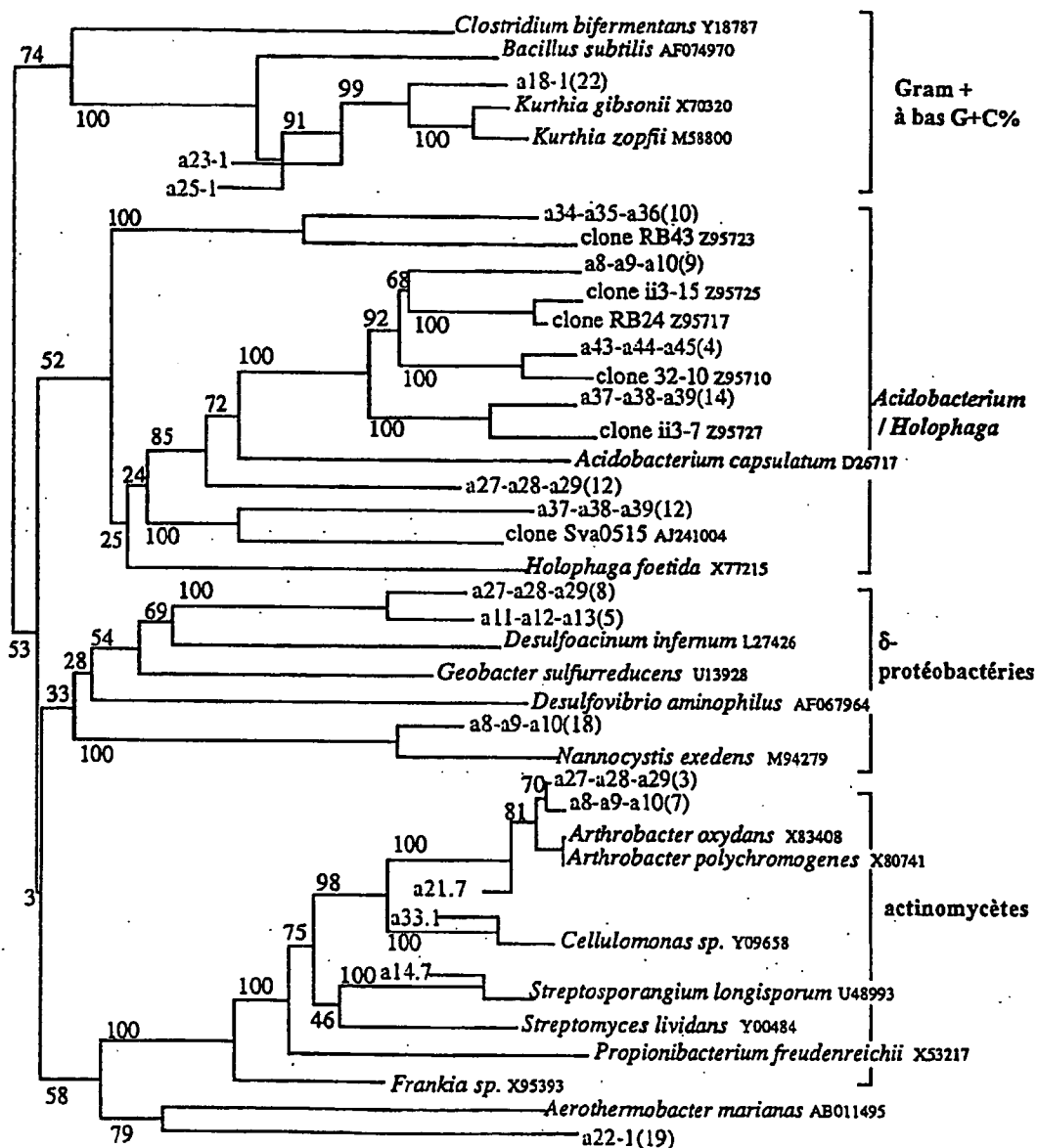


Figure 7- a)

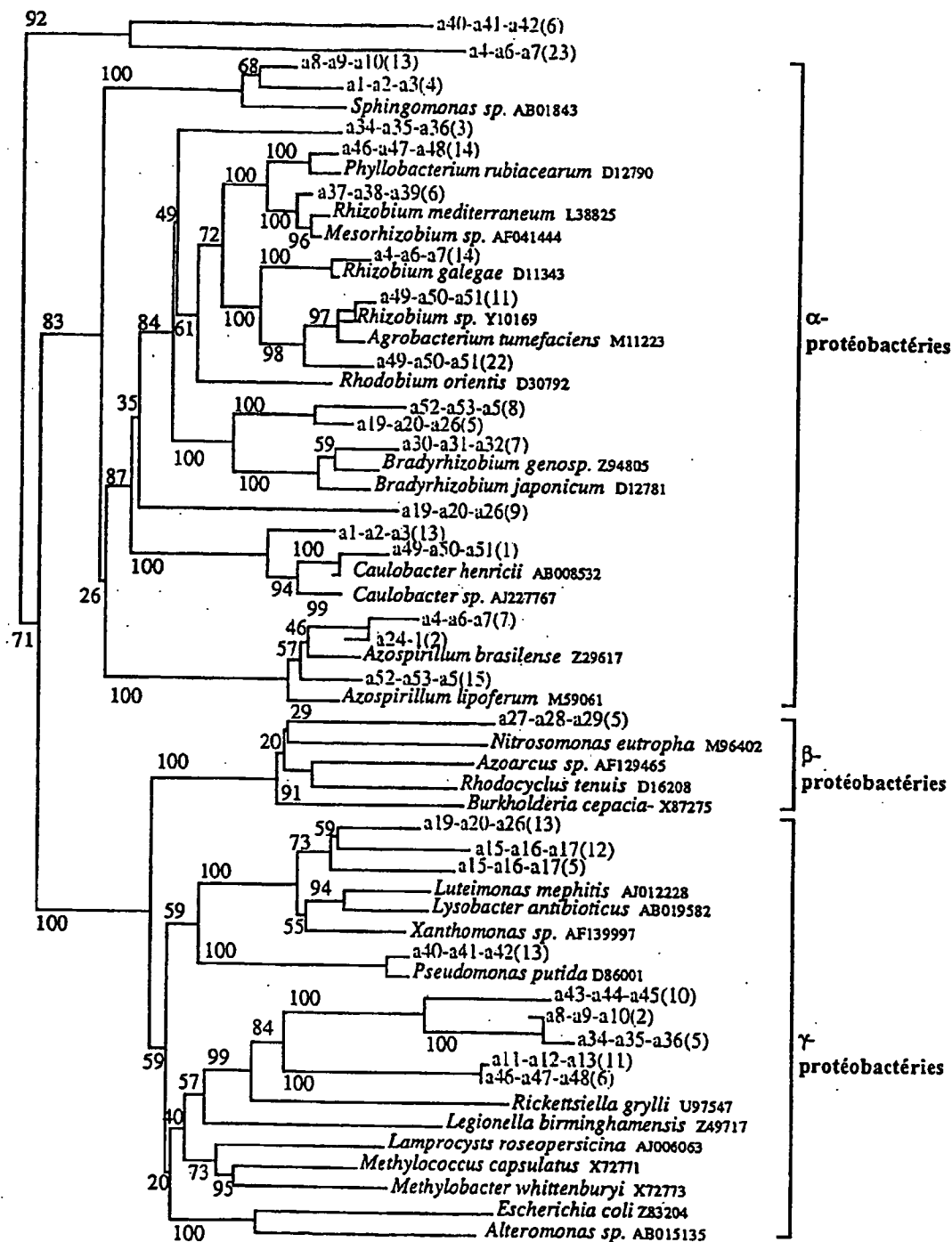


Figure 7 - b)

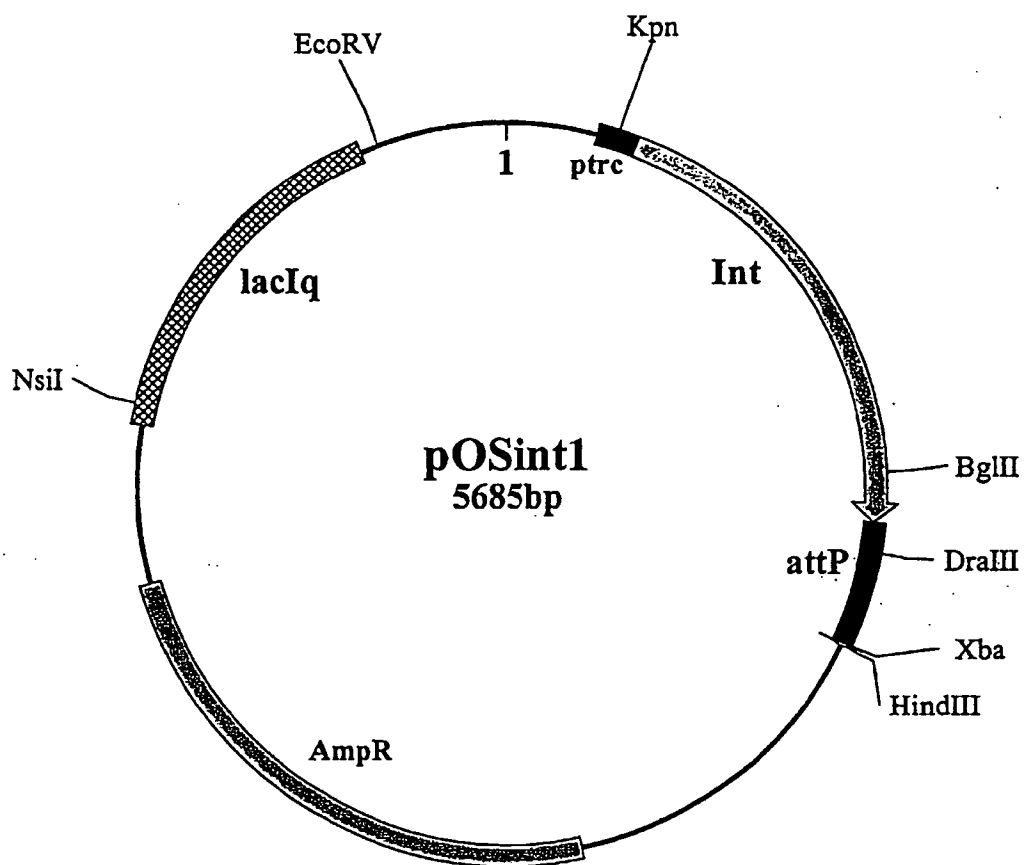


Figure 8

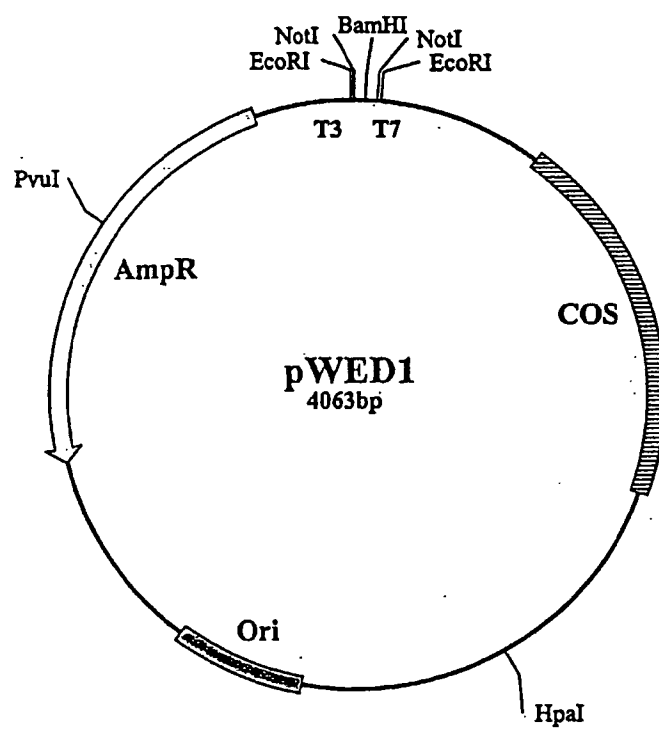


FIGURE 9

11/38

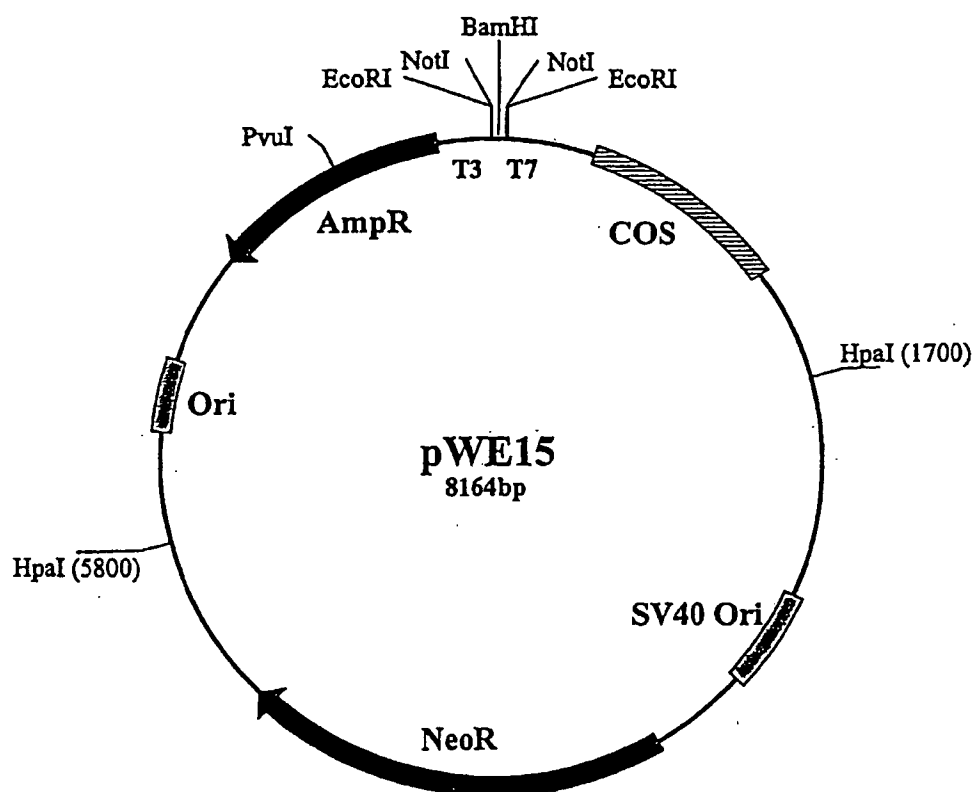


Figure 10

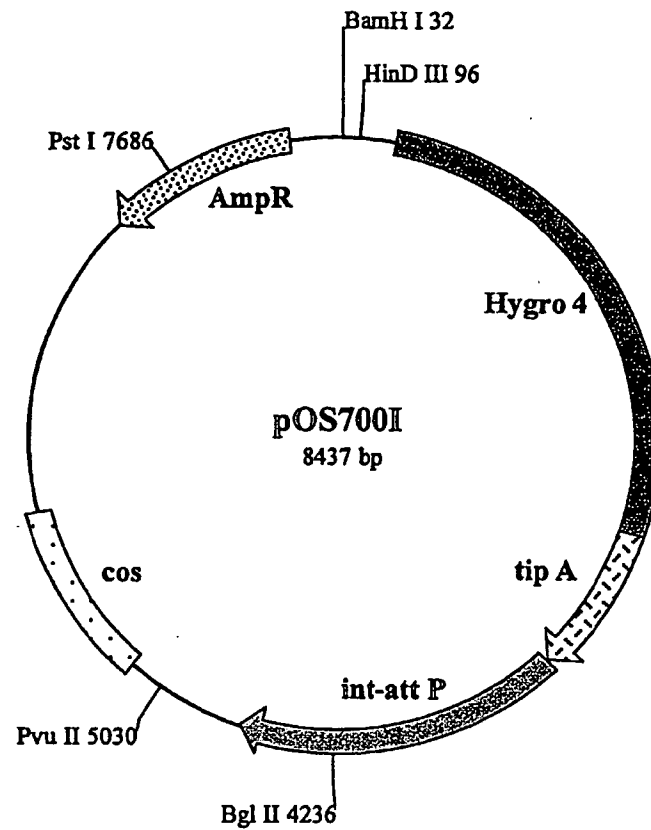


Figure 11

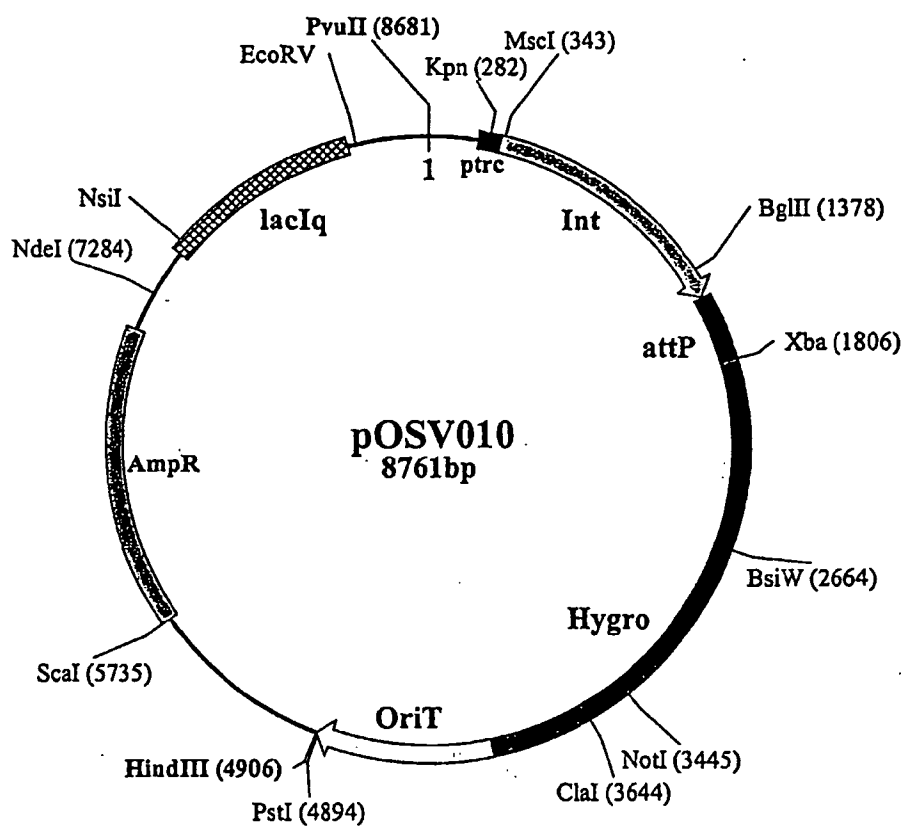
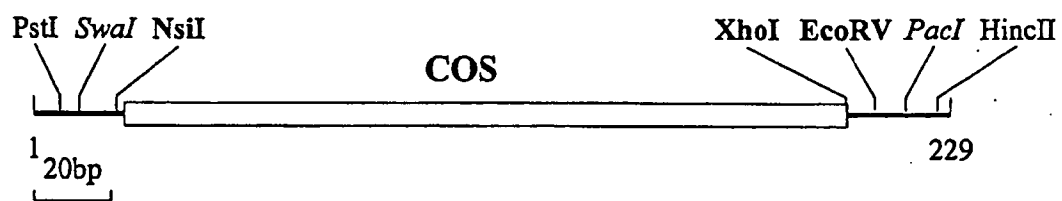


Figure 12

14/38



PCR COS

Figure 13

15/38

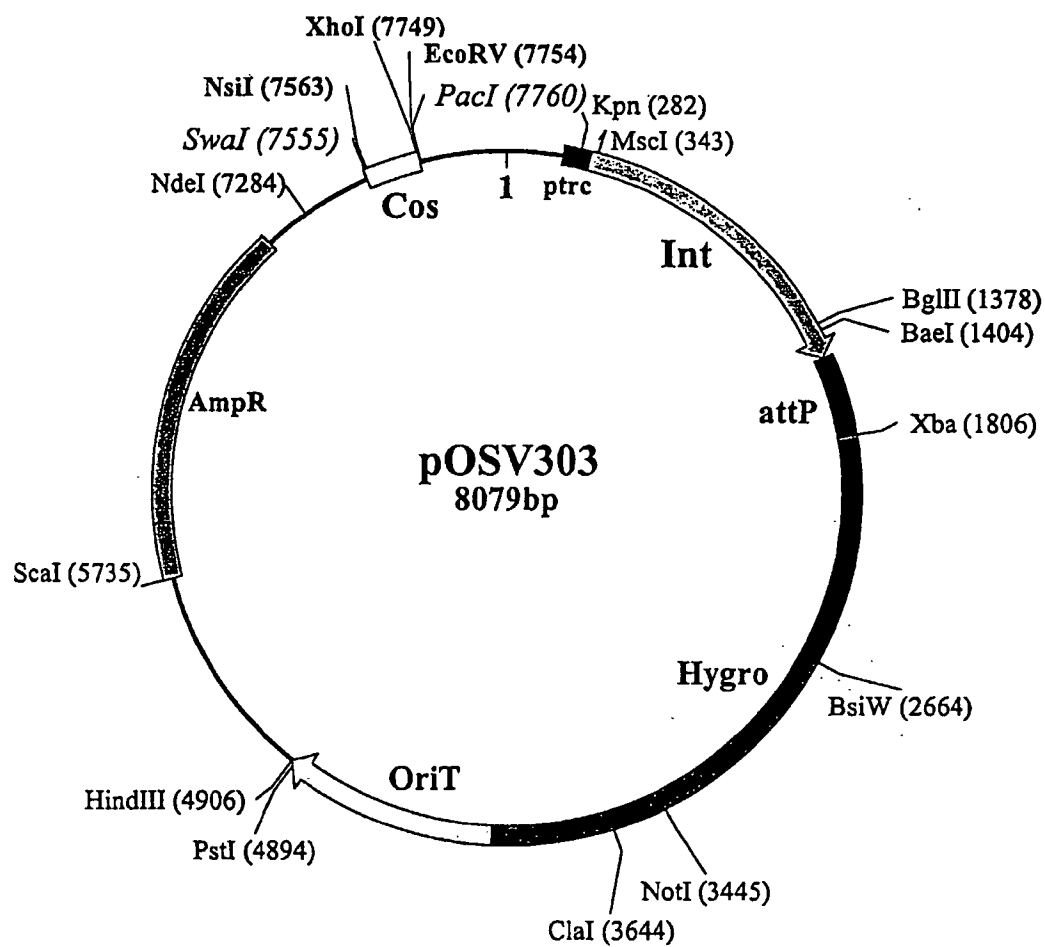


Figure 14

16/38

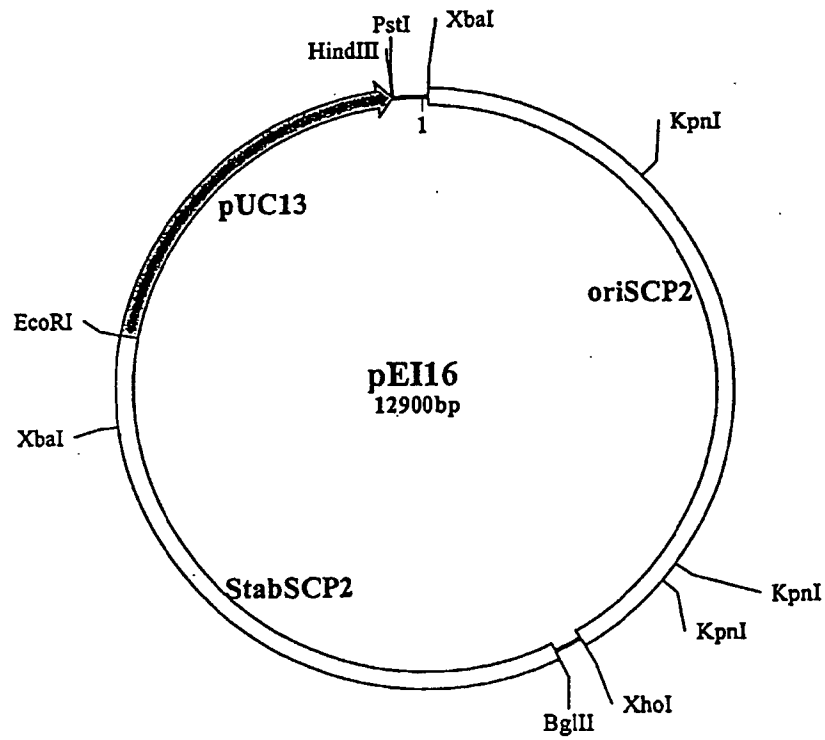


FIGURE 15

17/38

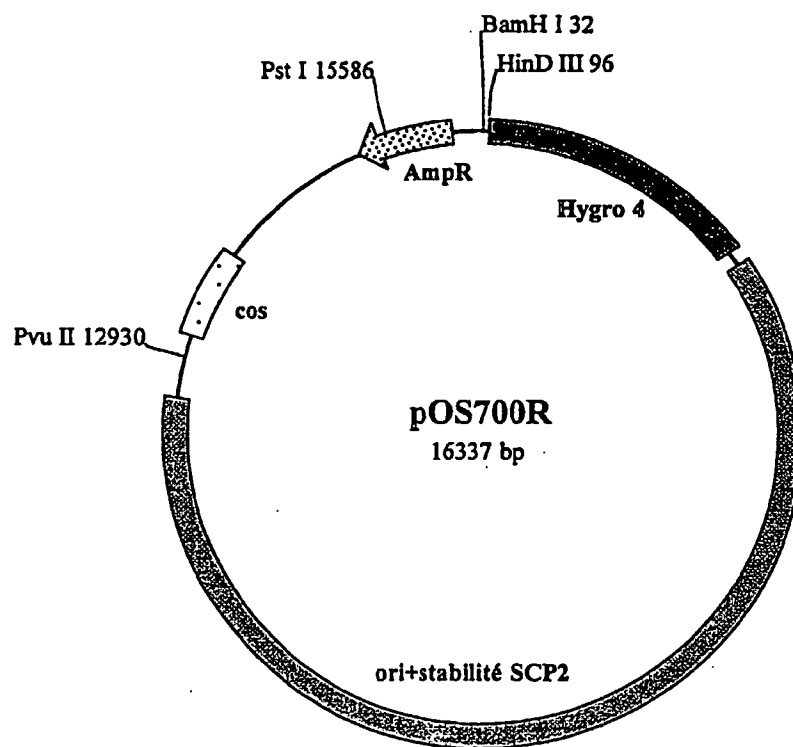


Figure 16

18/38

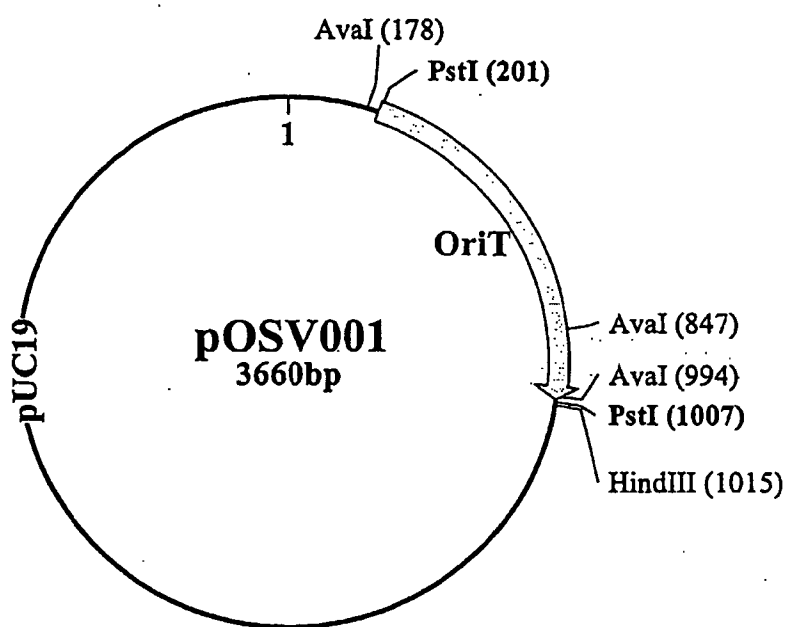


Figure 17

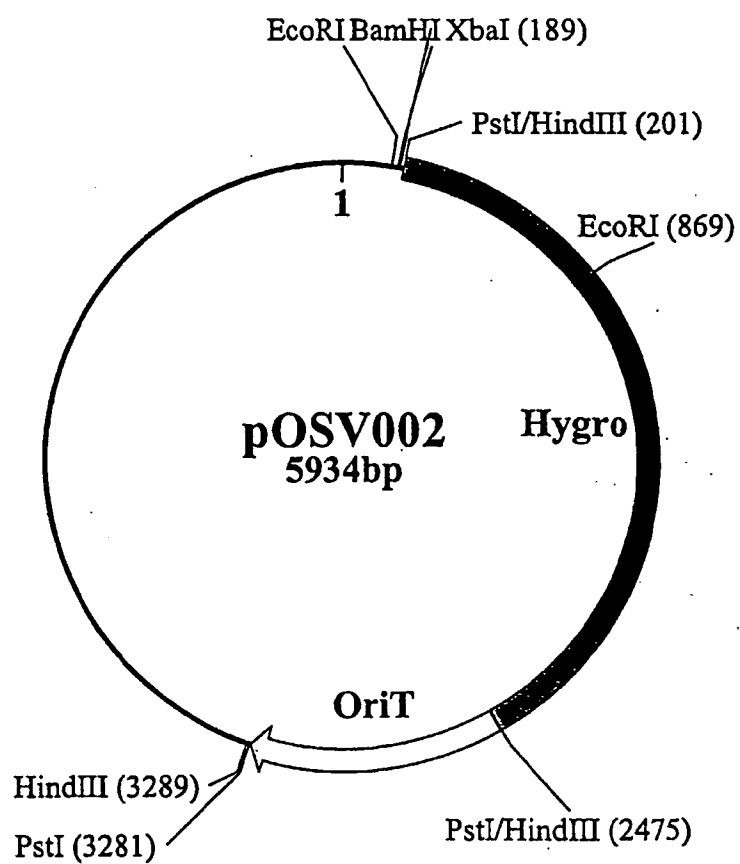


Figure 18

20/38

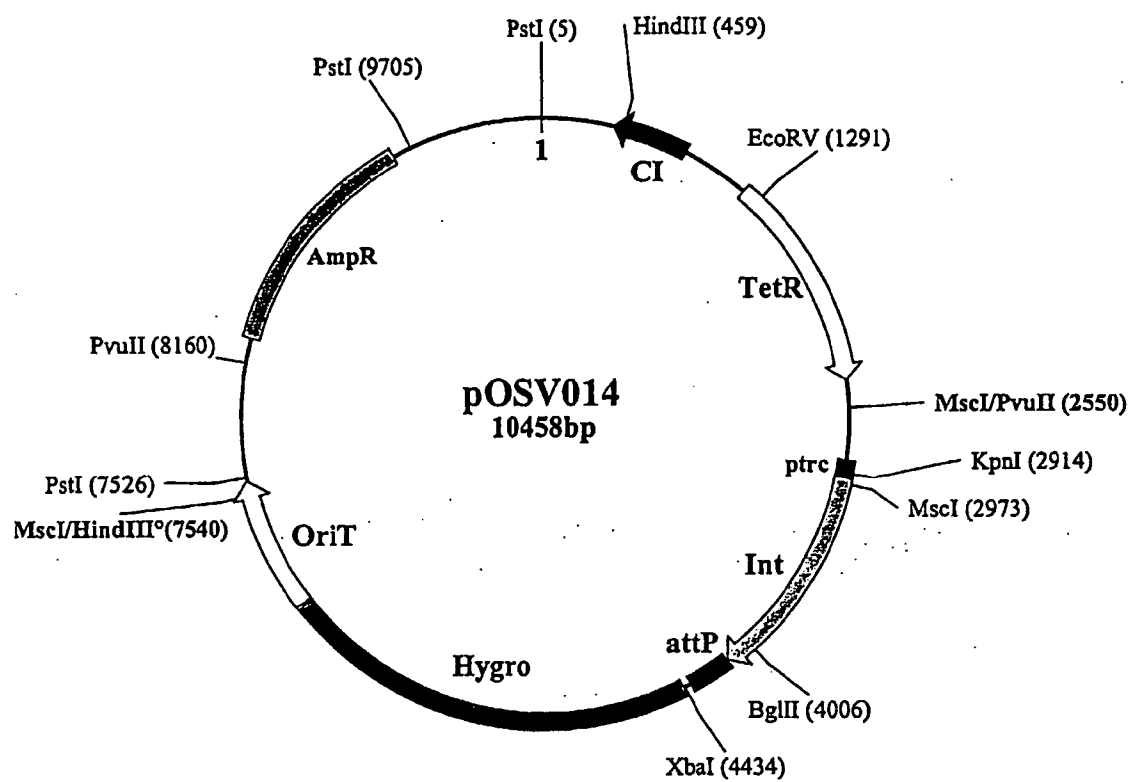


Figure 19

21/38

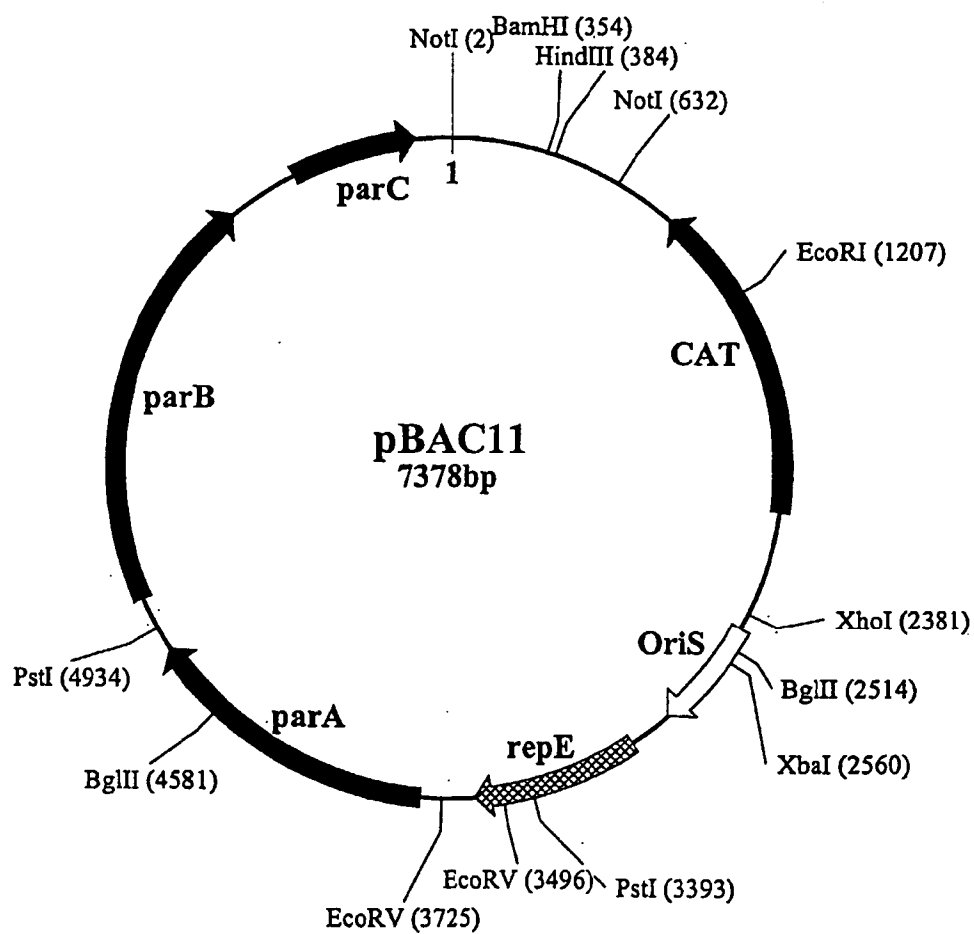


Figure 20

22/38

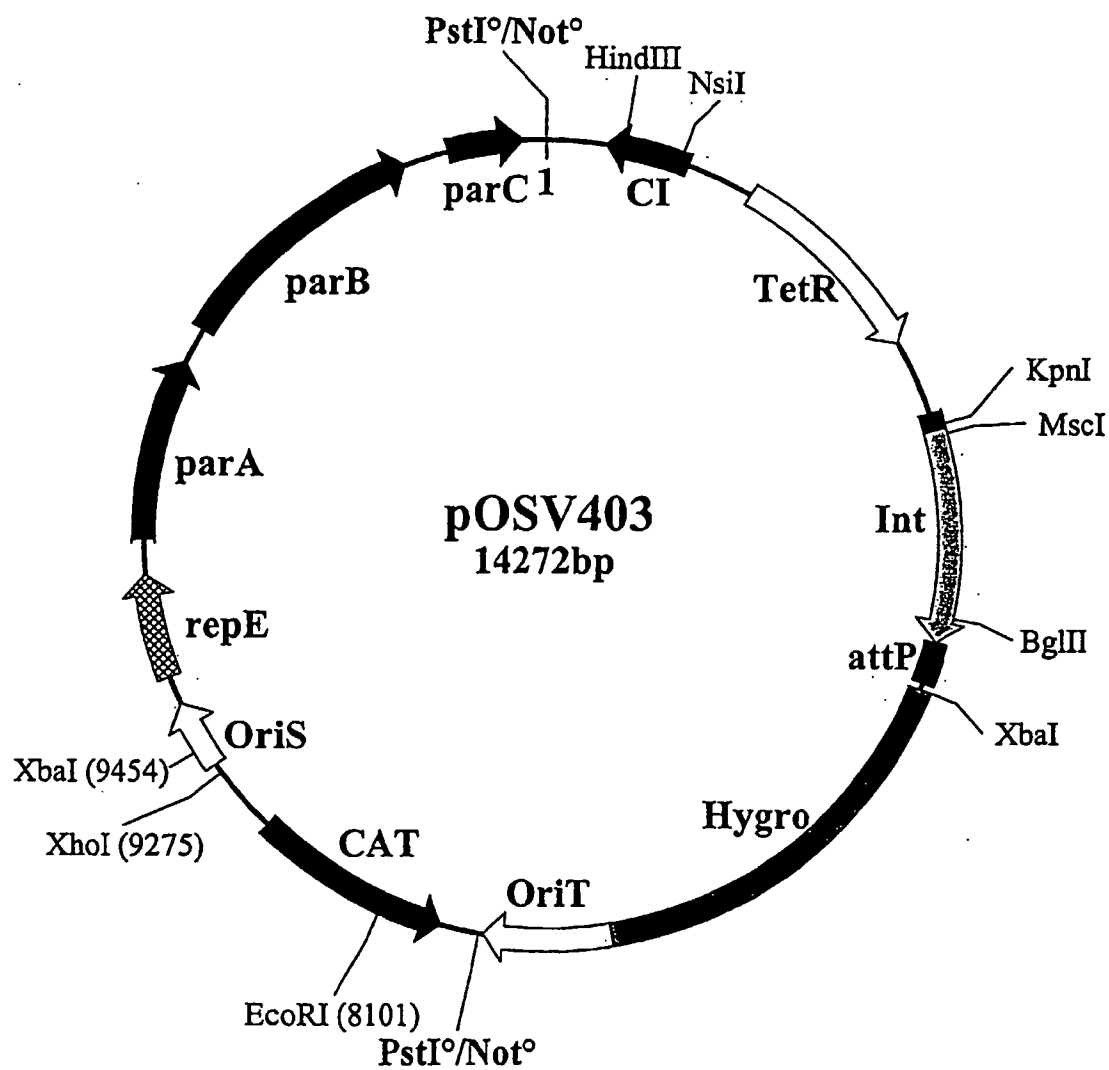


Figure 21

23/38

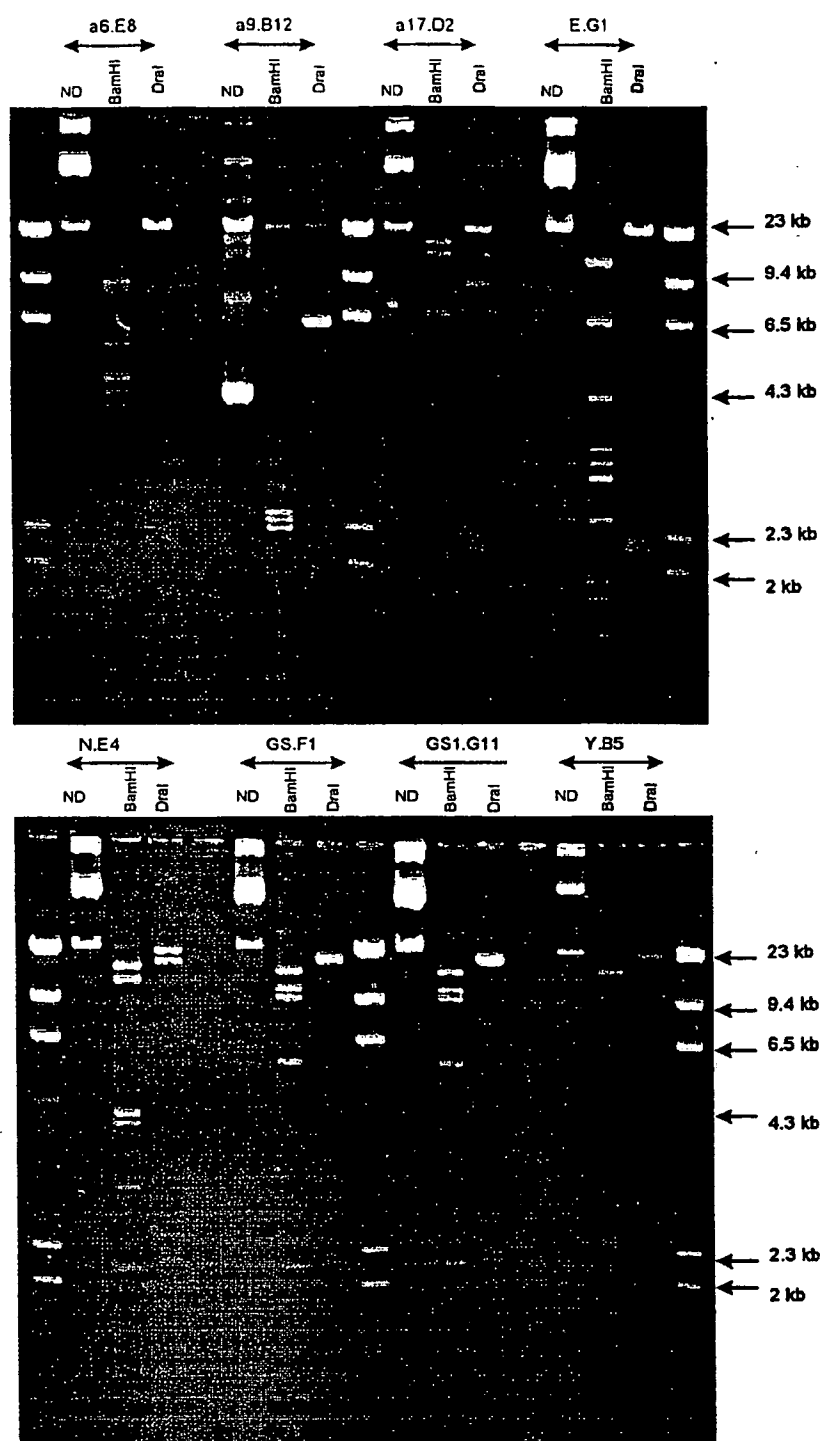


Figure 22

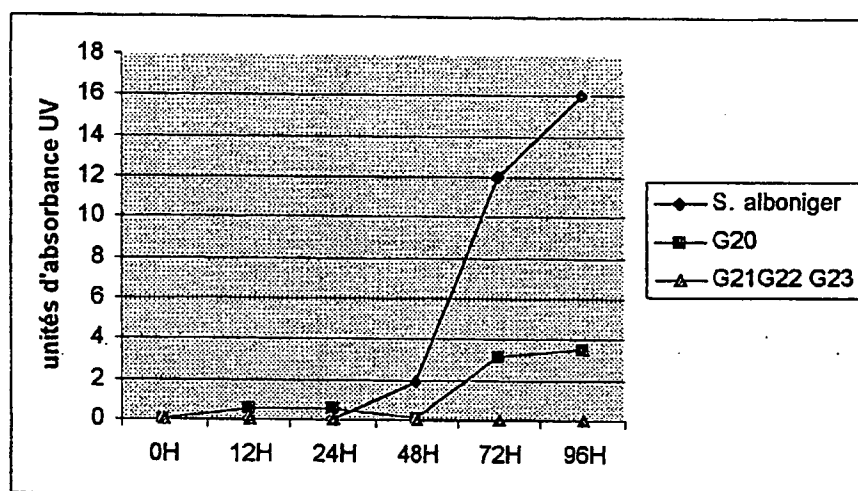


Figure 23

101					150
sol_a26G1-1	KDFLATRVSY	KLNLRGPSLT	VQTACSTSLV	SVVMACESLQ	RGASDIALAG
sol_a46B5-31	KDYLPTRVSY	KLNLRGPSLA	VQSACSTGLV	AVCQAIQNLO	TYQCDMALAG
sol_a9B12-3	KDFIATRtay	KLNLRGPAMA	VGtACSTSLV	AVHEACQALR	LGECDMALAG
sol_a49F1-32	KDFIATRtay	KLNLRGPAMT	VQTACSSSLV	AVHVAAQSLL	AGECDIALAG
<i>B. subtilis</i>	SGTIPTMISH	KLGLRGPSYF	VHANCSSSLI	GLHSAYKSLL	SGESDYALVG
stramb12	GSVLSGRIAY	TFGLQGPAVT	VDtACSSSLV	ALHLAAQALP	AGECELALVG
stramb9	AAVLSGRVSY	AFGLEGPAVT	VDtACSSSLV	ALHLAAQALR	RGECDLALAG
EryA (module 1)	TSVASGRIAY	TLGLEGPAIS	VDtACSSSLV	AVHLACQSLR	RGESSLAMAG
sol_a26G1-2	FSTAAGRISY	LLGLQGPNEP	VDtACSSSLV	AVHLACRSIQ	SRECSMALAG
sol_a53F11-13	LNAAAAGRLSY	VLGLQGPSMA	VDtACPSSLV	AIHLACQSLR	NRECRMALAG
sol_a53F11-14	HSIAAGRILAY	VLGLQGPAMA	VDtACSSSLV	AIHLACQSLR	NDDCRVAVAG
sol_a26G1-10	HSMLANRISY	LLDLRGPSMA	VDtACSSALV	AVHLACQSLR	RRECDAAEFAG
sol_a36E8-1	LSIAANRLSY	TFDFRGPSLA	VDtACSSSLV	AIHLACQSVR	RGEAE LAVAA
<i>M. tuberculosis</i>	MSIIANRLSY	FLDLRGPSVA	VDtACSSSLV	AIHLACQSLR	TQDCHLAIAA
sacery19	LSIIPARIAY	FLGLRGPDMT	LNTACSSALV	AMHQARQSIL	LGESSVALVG
sol_a17D2-3	LAVVANRISY	IYDLRGPSLT	VDtACSSSLV	ALHQAVEALR	SGRIETAIVG
fas humaine	RAMMANRLSF	FFDFRGPSIA	LDTACSSSLM	ALQNAVQAIH	SGQCPAAIVG

Figure 24

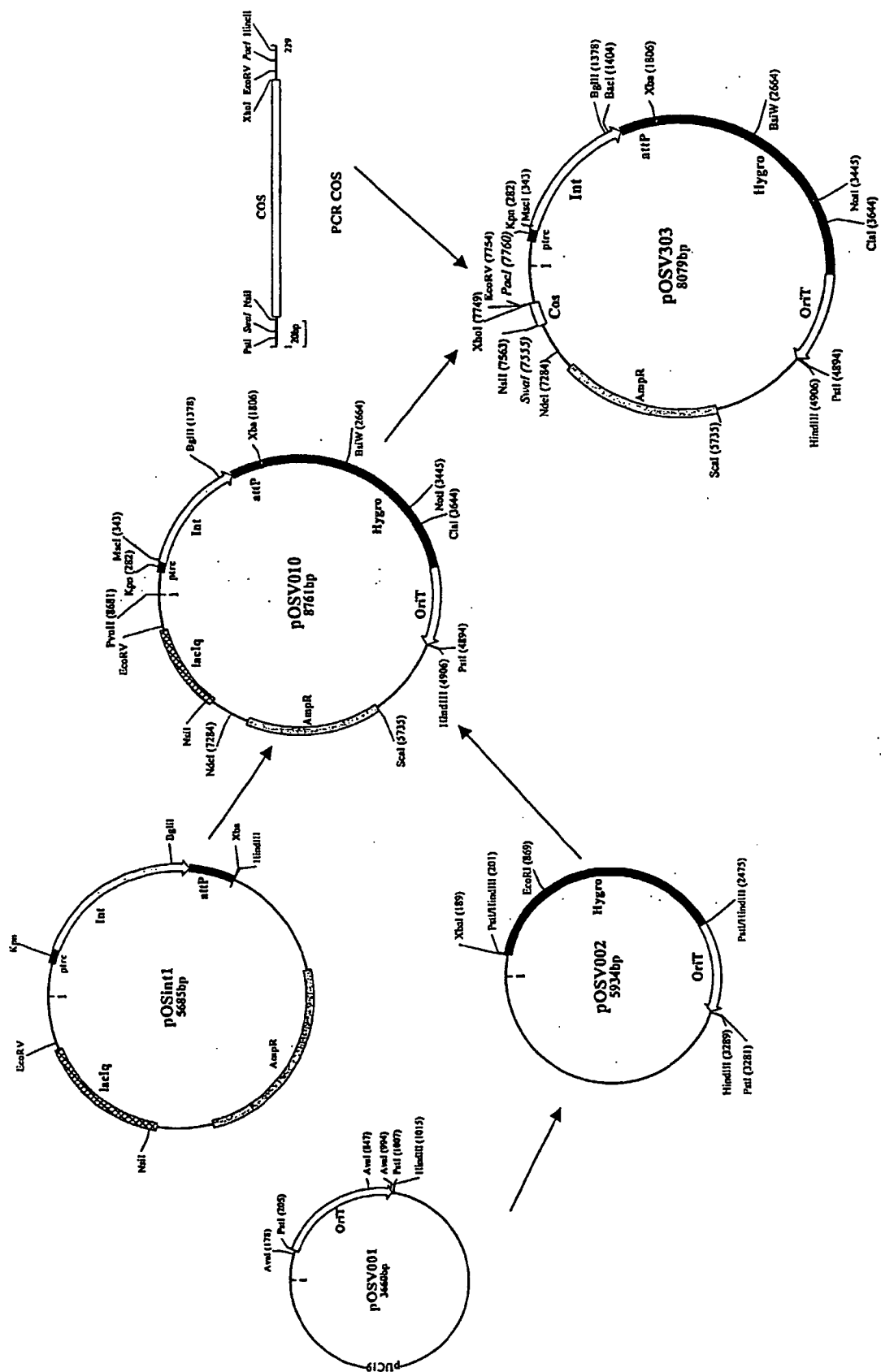


FIGURE 25

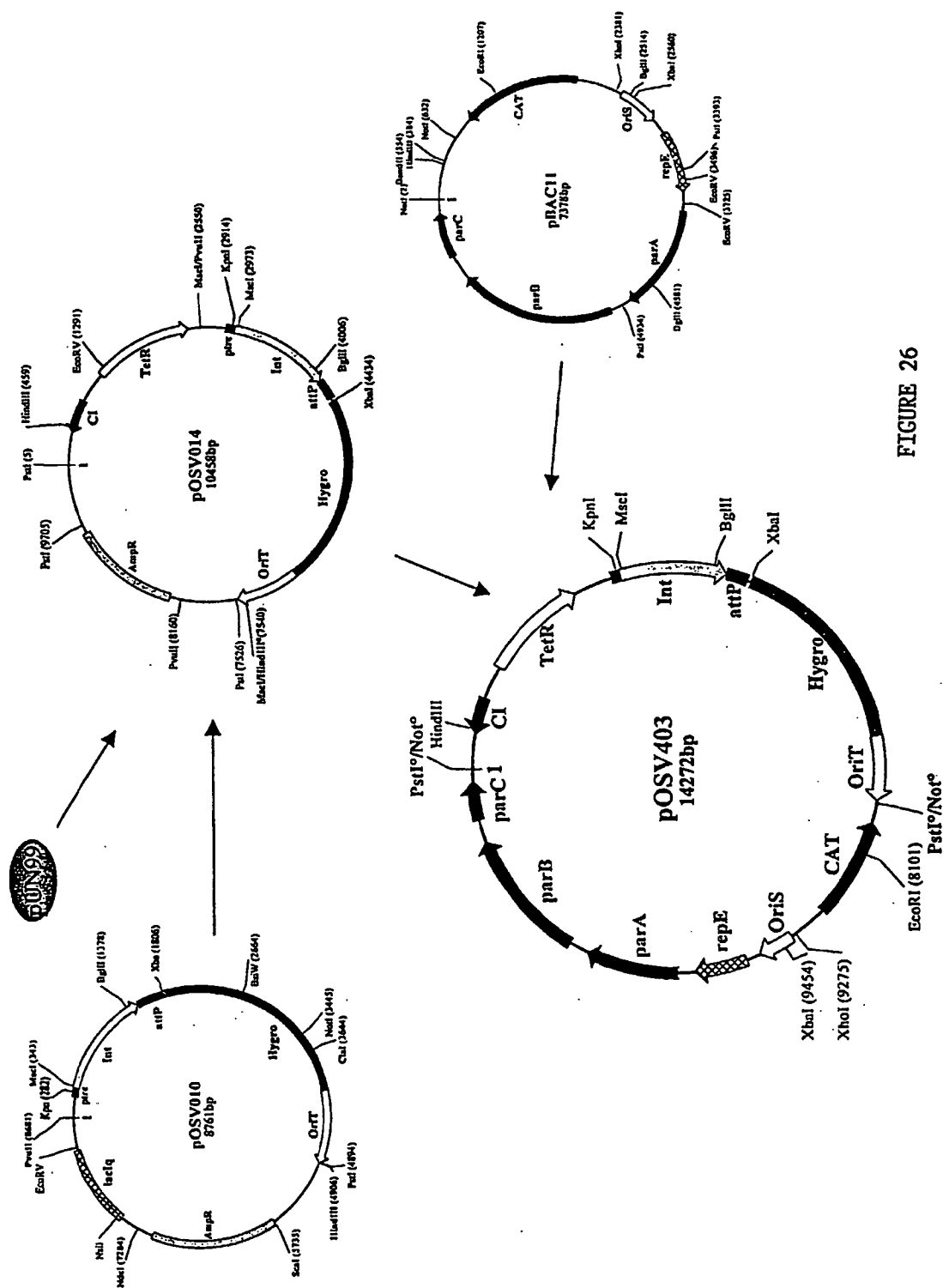


FIGURE 26

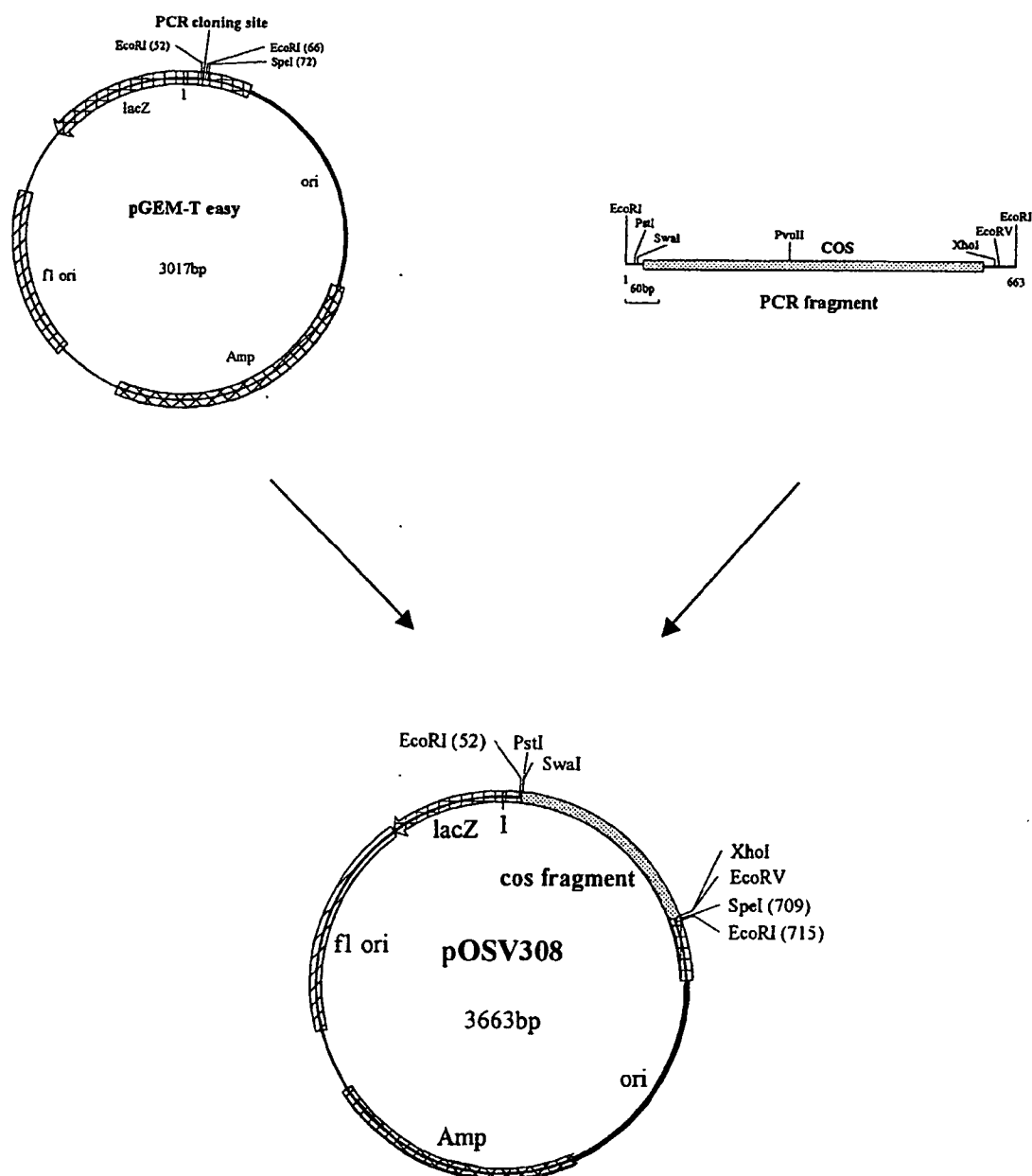


FIGURE 27

29/38

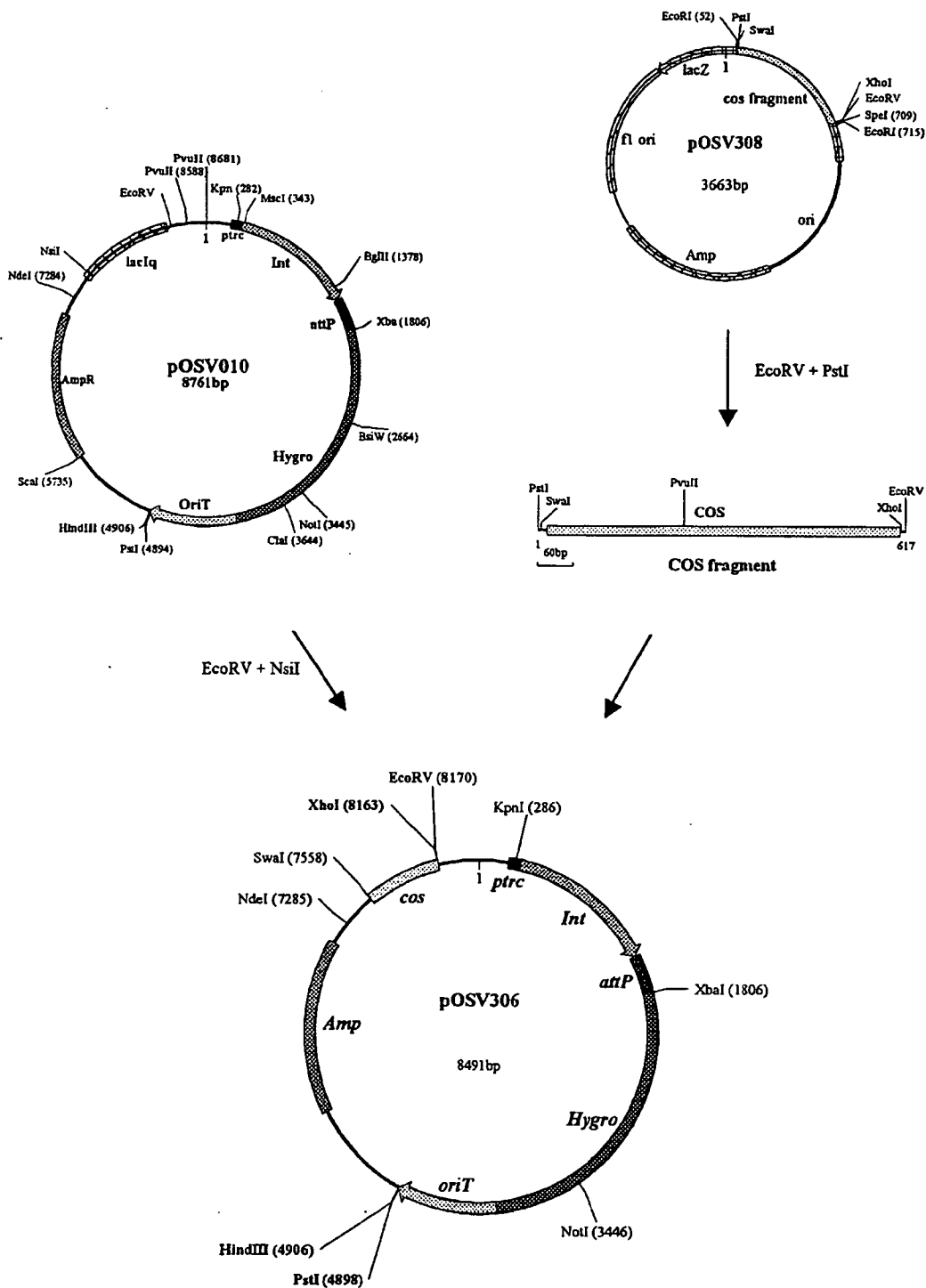


FIGURE 28

30/38

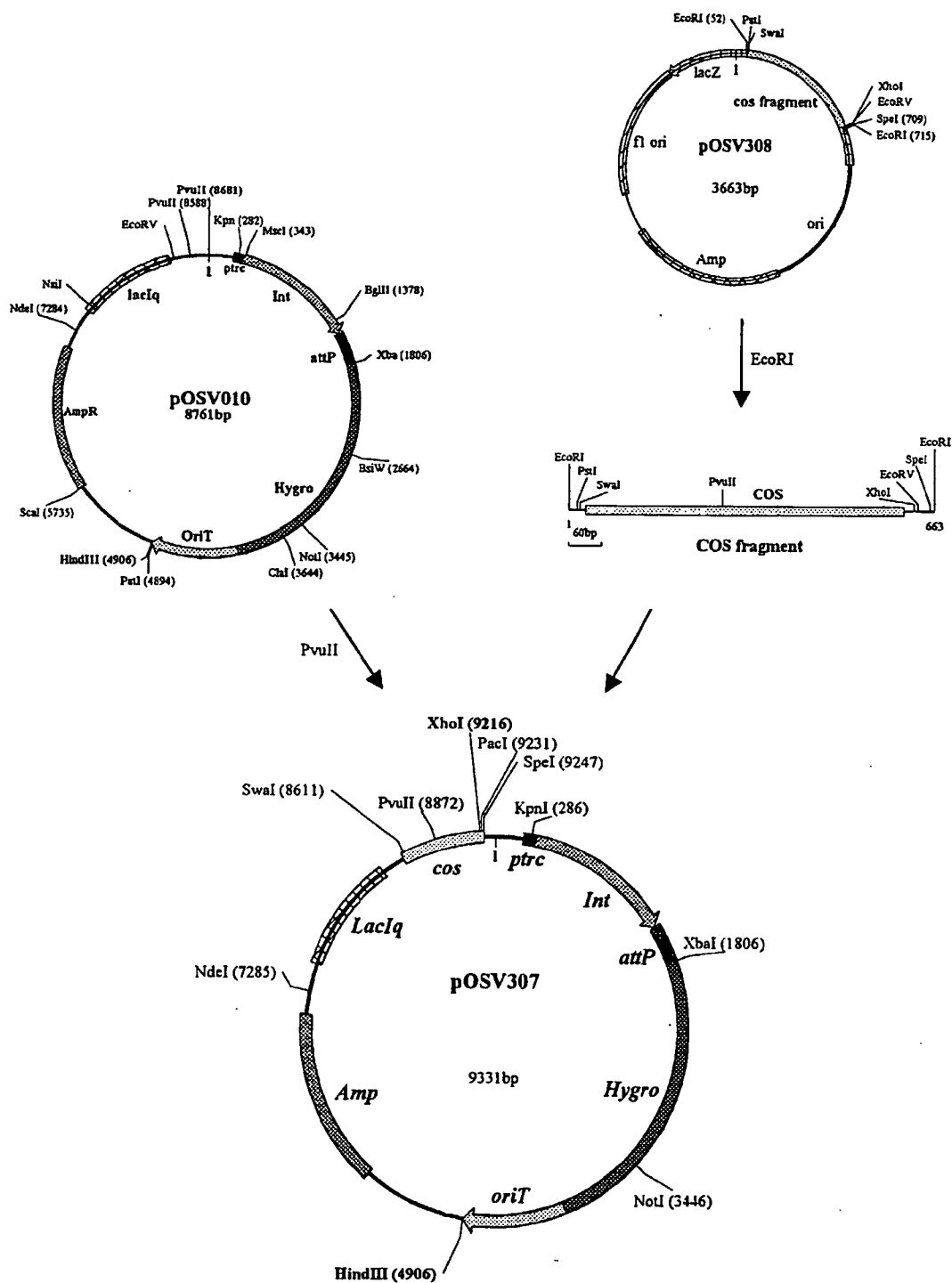


FIGURE 29

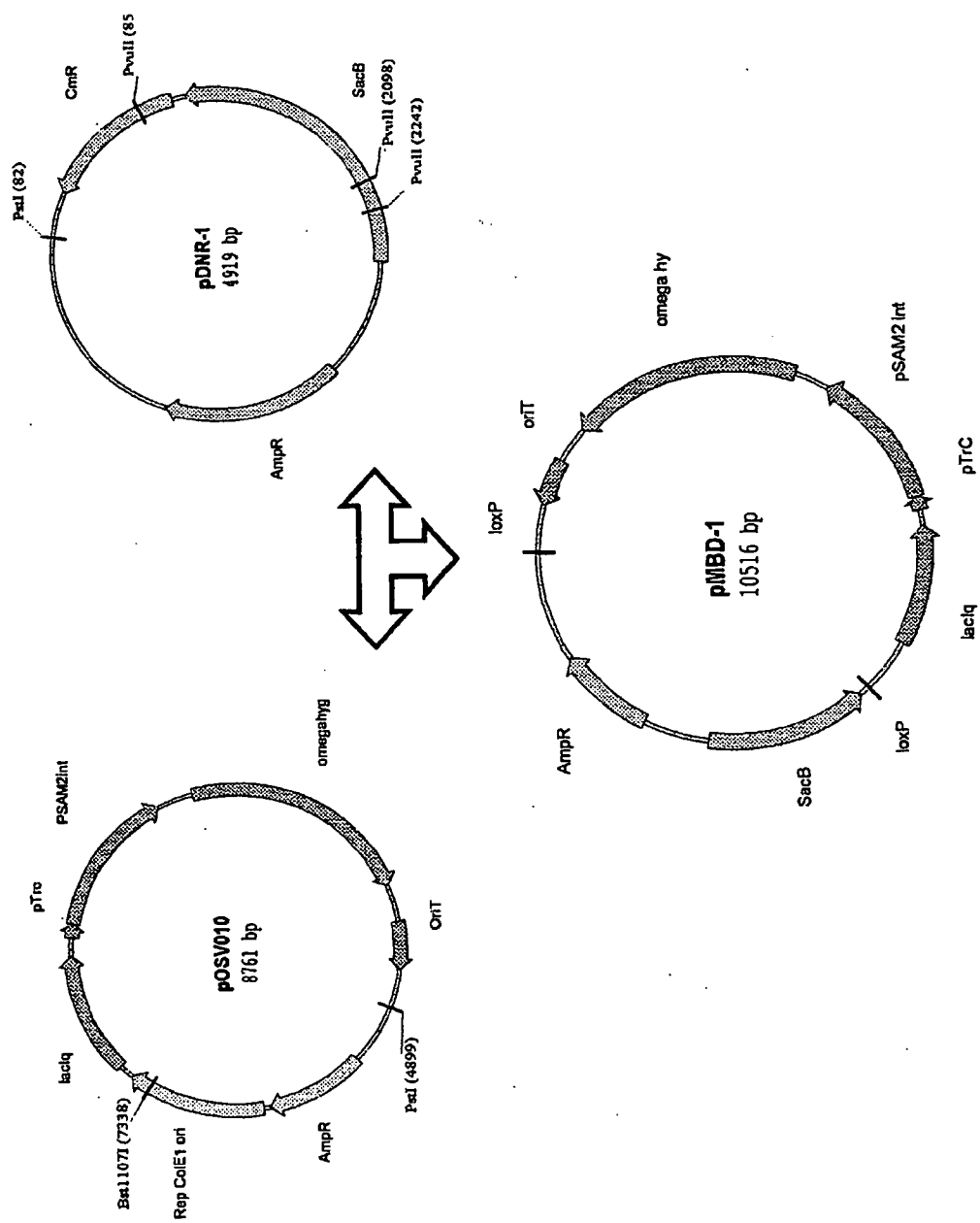


FIGURE 30

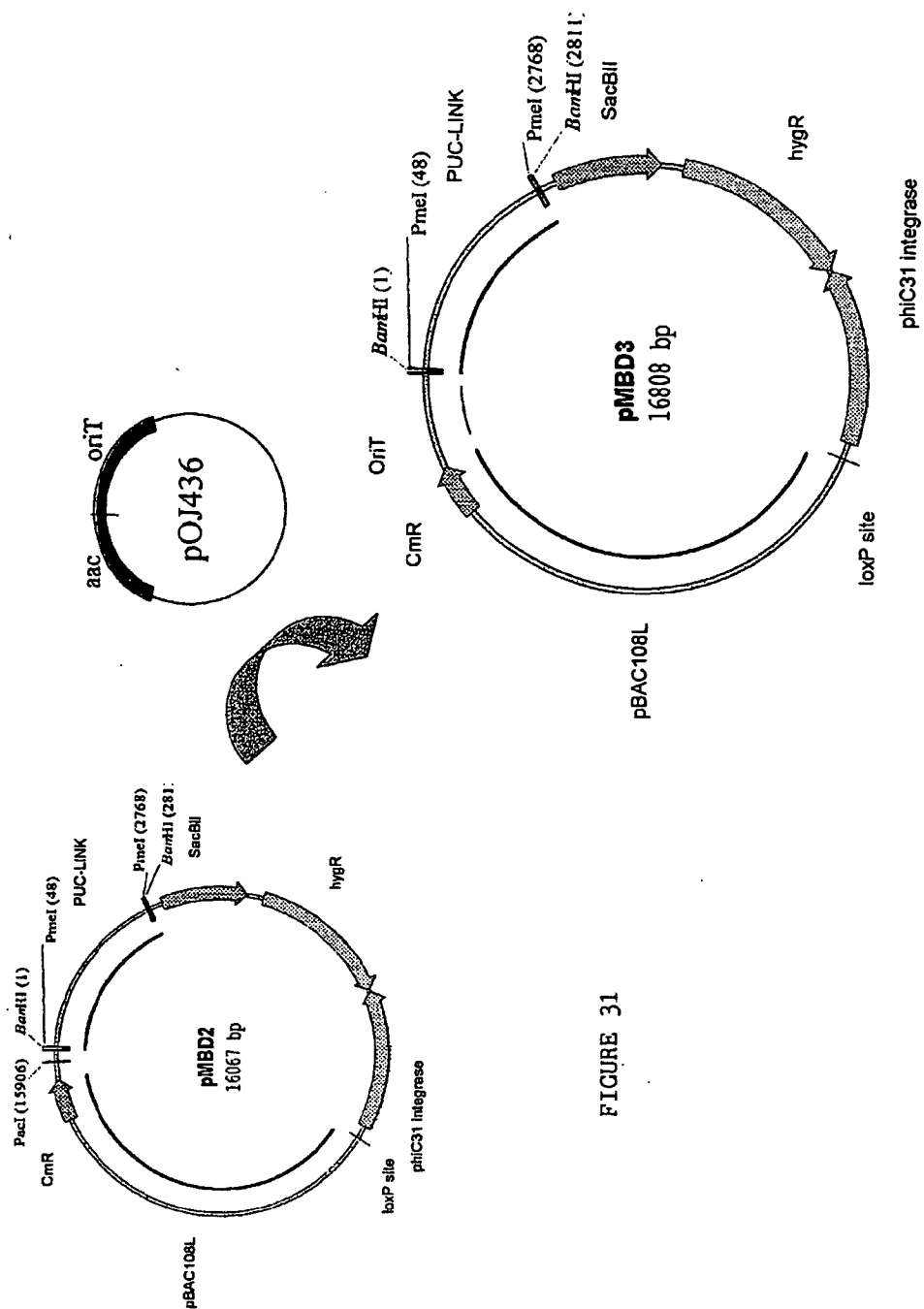


FIGURE 31

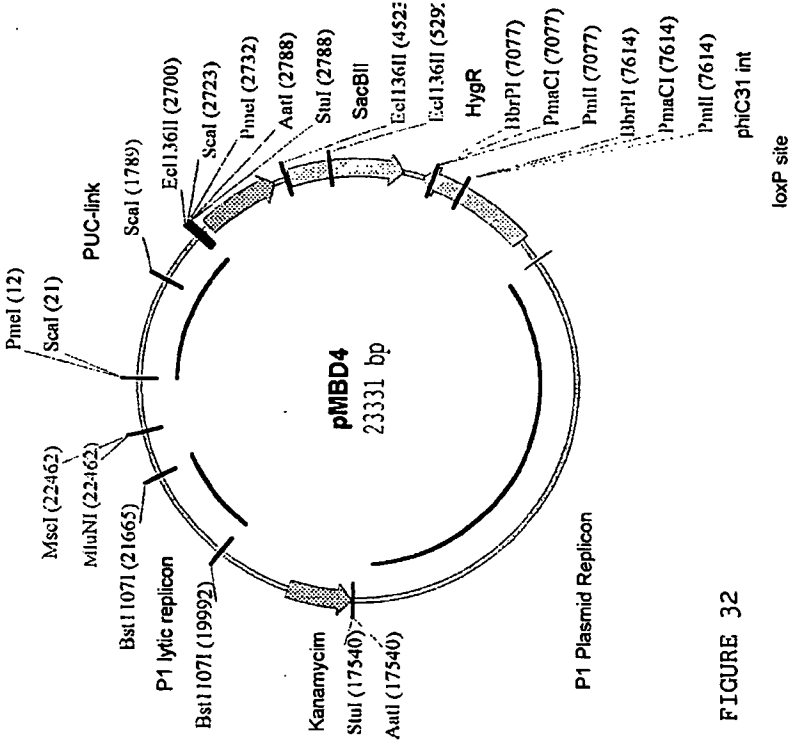


FIGURE 32

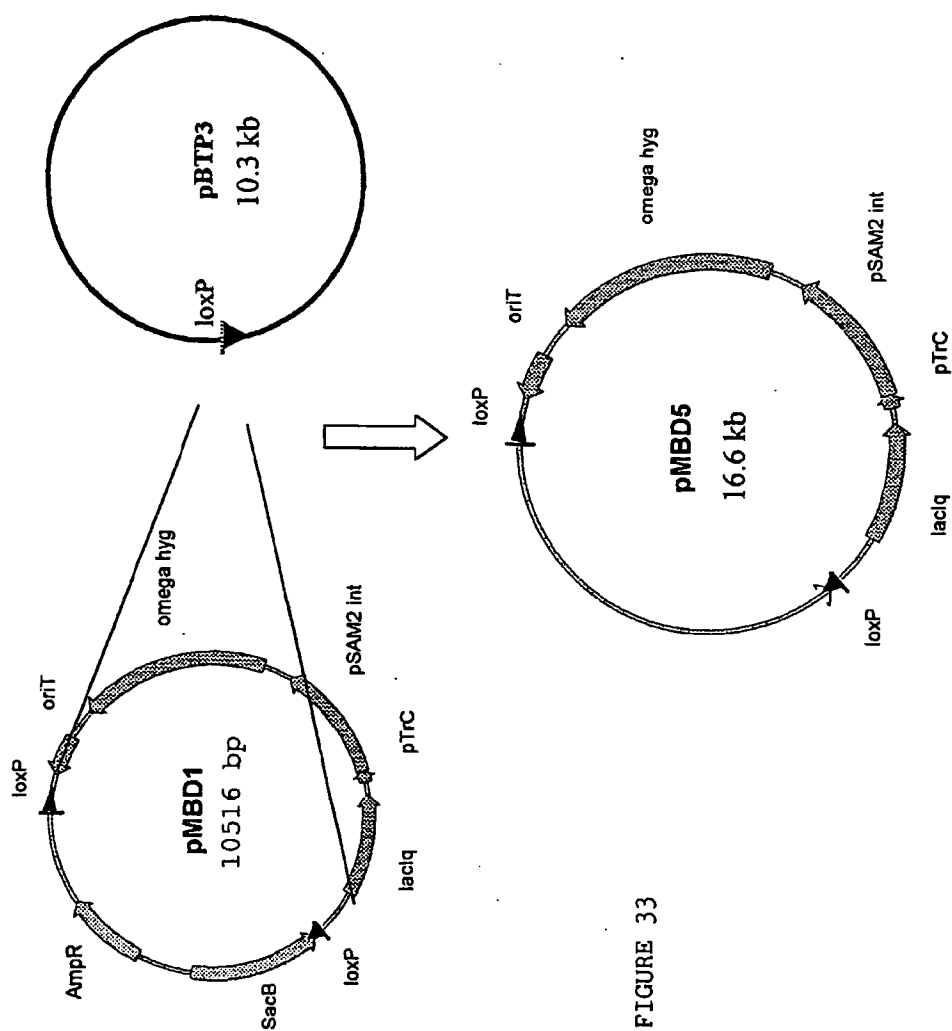


FIGURE 33

35/38

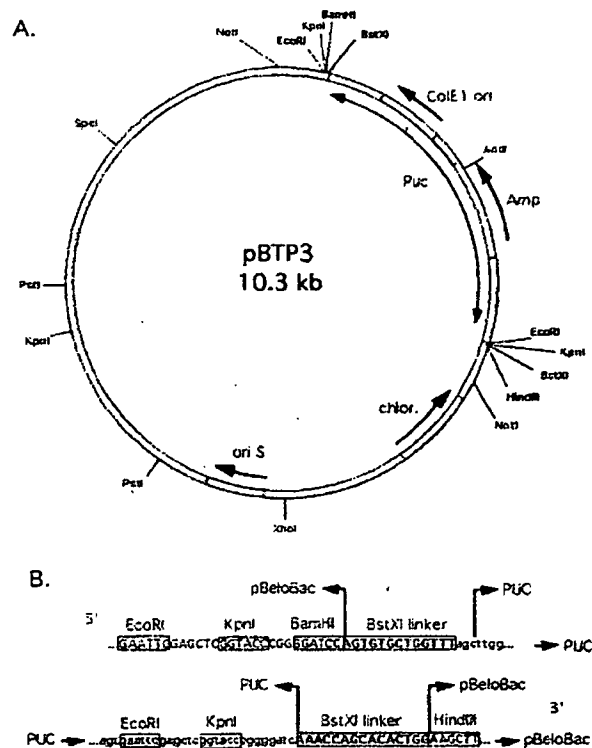


FIGURE 34

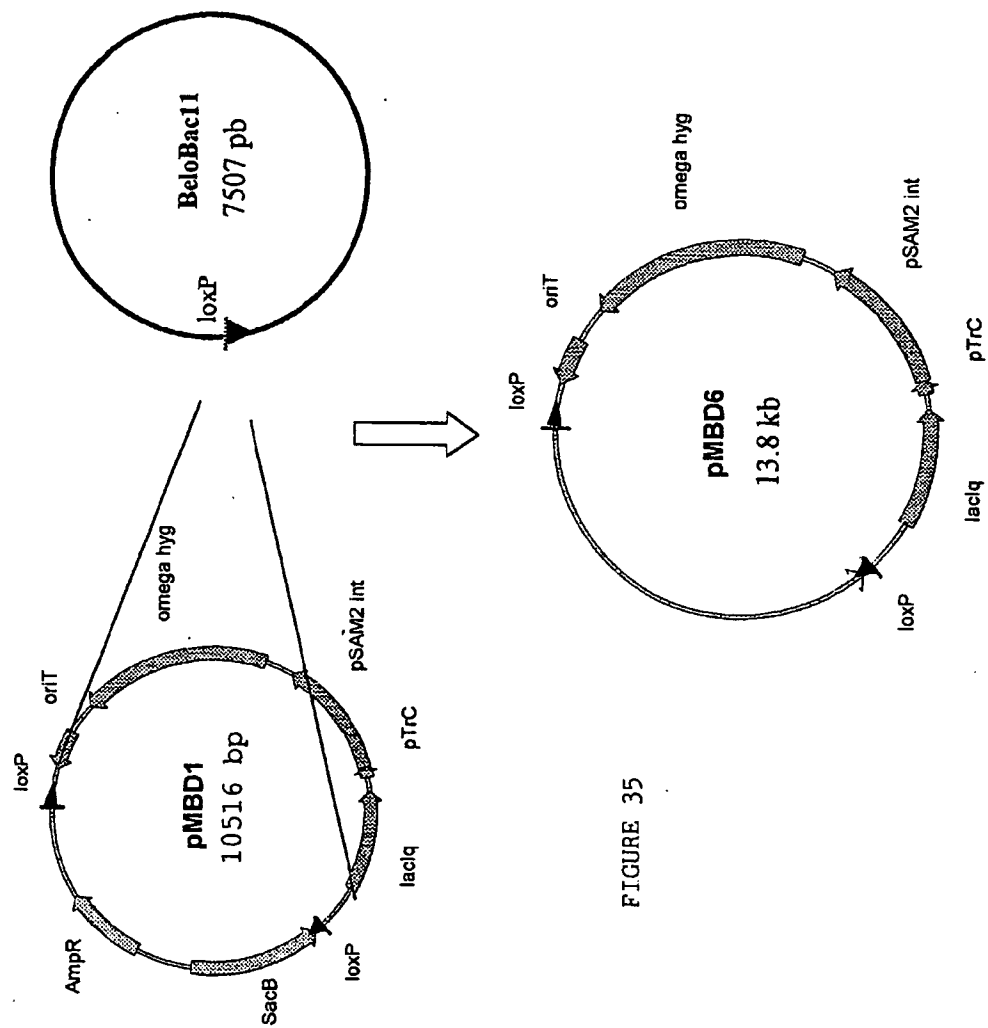


FIGURE 35

37/38

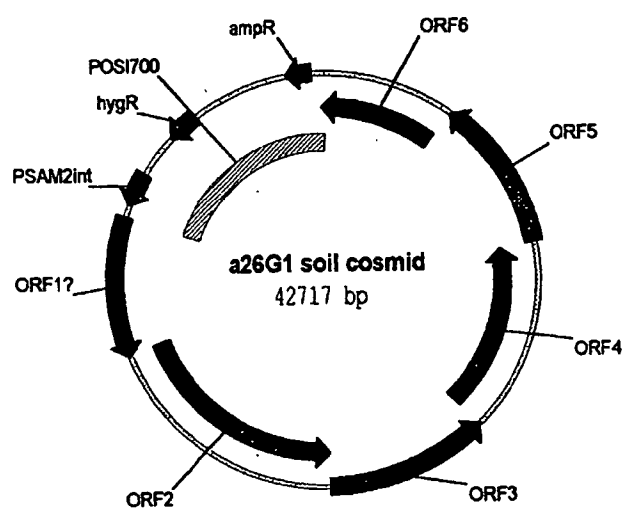


FIGURE 37

38/38

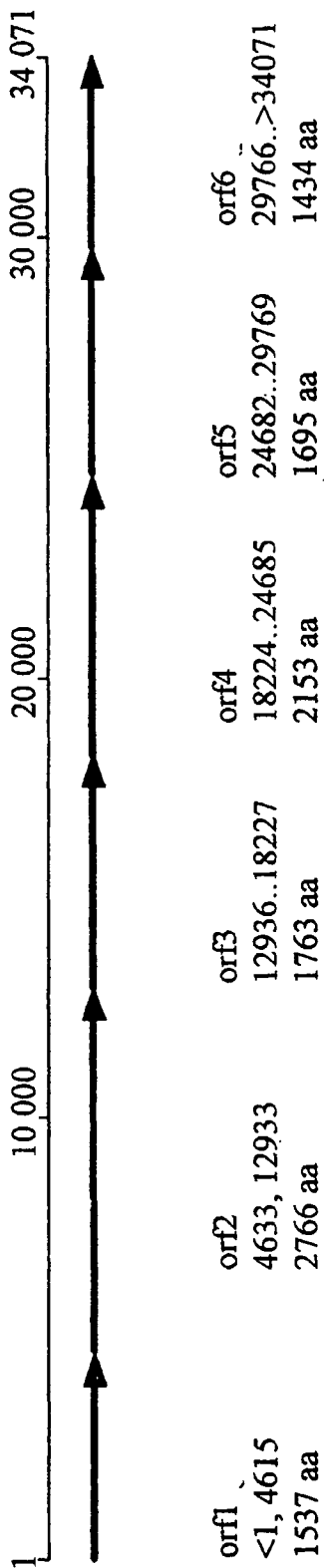


FIGURE 37

LISTE DE SEQUENCES

<110> Aventis Pharma S.A.

<120> Procédé d'obtention d'acides nucléiques à partir d'un échantillon de sol, acides nucléiques ainsi obtenus et leur application à la synthèse de composés

<130> Banque d'ADN du sol - RPR S.A.

<140>

<141>

<150> FR9915032

<151> 1999-11-29

<160> 126

<170> PatentIn Ver. 2.1

<210> 1

<211> 15

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:sonde
FGPS431

<220>

<221> variation

<222> (14)

<223> Base A remplacée par G

<400> 1

acgggcggtg tgtac

15

<210> 2

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce
FGPS122

<400> 2

ggagagtttg atcatggctc ag

22

<210> 3
<211> 20
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:amorce
FGPS350

<400> 3
cctggagtta agccccaagc

20

<210> 4
<211> 24
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:sonde
FGPS643

<220>
<221> variation
<222> (20)
<223> T remplacée par C

<400> 4
gtgagtnnna acctgcccct gact

24

<210> 5
<211> 21
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:sonde
FGPS643-2

<400> 5
gtgagtaacc tgcccccgac t

21

<210> 6
<211> 23
<212> ADN
<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce
R499

<400> 6

ttaattcact tgcaactgat ggg

23

<210> 7

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce
R500

<400> 7

aacgatagct cctacatttg gag

23

<210> 8

<211> 25

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:sonde C501

<400> 8

ttgctgatac ggtatagaac ctggc

25

<210> 9

<211> 20

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce
FGPS516

<400> 9

tccagatcct tgacccgcag

20

<210> 10

<211> 20

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce
FGPS517

<400> 10

cacgacattg cactccaccg

20

<210> 11

<211> 16

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:sonde
FGPS518

<400> 11

ccgtgagccg gatcag

16

<210> 12

<211> 20

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:FGPS612

<220>

<221> variation

<222> (2)

<223> Base C remplacée par T

<220>

<221> variation

<222> (7)

<223> Base T remplacée par C

<220>

<221> variation

<222> (7)

<223> Base T remplacée par A

<400> 12

ccaacttcgt gccagcagcc

20

<210> 13

<211> 21
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS669

<220>
<221> variation
<222> (7)
<223> Base A remplacée par G

<220>
<221> variation
<222> (13)
<223> Base A remplacée par C

<400> 13
gacgtcatcc ccaccttcct c 21

<210> 14
<211> 18
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS618

<220>
<221> variation
<222> (5)
<223> Base T remplacée par C

<400> 14
atggttgctg tcagctcg 18

<210> 15
<211> 21
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS614

<400> 15
gtgtagaagt gaaattcgat t 21

<210> 16
<211> 18
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS615

<400> 16
cgggtggatga tgtggatt 18

<210> 17
<211> 18
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS616

<400> 17
aggttaaaac tcaaatga 18

<210> 18
<211> 18
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS621

<400> 18
atacgtaggt ggcaagcg 18

<210> 19
<211> 19
<212> ADN
<213> Séquence artificielle

<220>
<223> Description de la séquence artificielle:FGPS617

<400> 19
gccgggggtca actcggagg 19

<210> 20
<211> 18

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:FGPS680

<220>

<221> variation

<222> (11)

<223> Base A remplacée par C

<220>

<221> variation

<222> (11)

<223> Base A remplacée par T

<220>

<221> variation

<222> (13)

<223> Base T remplacée par A

<400> 20

tgagtcccca actccccg

18

<210> 21

<211> 20

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:FGPS619

<400> 21

gcttggggct taactccagg

20

<210> 22

<211> 21

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce 63f

<400> 22

caggcctaac acatgcaagt c

21

<210> 23

<211> 18

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce
1387r

<400> 23

gggcggngtg tacaaggc

18

<210> 24

<211> 30

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:oligo-1

<400> 24

gcttatttaa atattaagcg gccgcccggg

30

<210> 25

<211> 28

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:oligo-2

<400> 25

cccgggcggc cgcattaata tttaaata

28

<210> 26

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce a1

<400> 26

ccncagnagc gcntnttntct nga

23

<210> 27

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce a2

<400> 27

gtncncgtnc cgtgngtntc na

22

<210> 28

<211> 23

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce b1

<400> 28

ccncagnagc gcntnctnct nga

23

<210> 29

<211> 22

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce b2

<400> 29

gtncncgtnc cgtgngcctc na

22

<210> 30

<211> 672

<212> ADN

<213> Streptomyces ambofaciens

<400> 30

ccccagcagc acgtgttcct cgagacggtg tgggagacct tcgaatccgc cggagtggac 60
 ccgcgcgcgg tacgcggtcg ttccgtcggg atgttcgtcg gcaccaacgg acaggactac 120
 ccggtggtgt tggccggatc cgccgacgag ggcctggacg cccacgcggc caccggtaac 180
 ggggcggcgg tgctgtccgg ccgggtctcg tacgccttcg gcctggaagg gccggcggtc 240
 accgtcgaca cggcgtgttc gtcgtcgctg gtggcccttc acctggccgc gcaggcgctg 300
 cggcgcggcg agtgcatct ggcactcgcc ggcggtgtgt cggagatgtc caccgaggcg 360
 gcgttcaccg agttcgccc gcagggcggc ctggccgacg acggccgctg caaggccttc 420
 tcggccgacg ccgacggcac gggctggggc gagggcgctg gcgtcctgct ggtggagcgg 480
 ctggcggacg cccgccgcaa cgggcaccgg gccctcgcgc tggtagggg cagcgcgggtc 540
 aaccaggacg gcgcctcaa cggctcgacg gcaccaacg gcccgctcca gcagcgagtc 600

atccggcagg cactggcgga cggccggctg tcgcccgcgg aggtcgacgc ggtcgagacc 660
 cacggcaccg gc 672

<210> 31
 <211> 665
 <212> ADN
 <213> Streptomyces ambofaciens

<400> 31
 ccccgagcgc gcgtgttcct ggaagcgtcc tgggagggcg tcgagcgggc aggcacgcac 60
 atgcgcaccc tcgcccgggtg acgcaccggc gtcttcgccg gcgtgatgta ccacgactac 120
 ccgtcgggtg tcgaccccga agcgtcgcac ggctacctgg gcacggccaa cggccggcagc 180
 gttctctccg gccgcatcgc ctacaccttc gggcttcagg gaccggcggg caccgtggac 240
 acggcctgct cctcgtccct ggtggcgctg cacctcgccg ccaggcgct gcccgcggc 300
 gagtgcaaac tcgccctggg cgggtggggtc acggtcatgt ccggcccgat gatgttcgcg 360
 ggcttcggcc tgggaagacgg ctctgcgcgc gacggccgct gcaaggcgtt cgccgcggcc 420
 gccgacggca ccggctgggg cgagggtgtc ggtgtgctgc tgggtggagcg gctgtcggac 480
 gcccggcgcc acgggcaccg ggtgctggcc gtggtgcgcg gtagcgcggt caaccaggac 540
 ggtgcctccg gcggcctcac cggccccaac ggacctgcc agcagcgcggt catccgtcag 600
 gccctggcga gcgcggcact cgtaccggcc gaggtcgacg cggtcgagac ccacggcacc 660
 gggac 665

<210> 32
 <211> 671
 <212> ADN
 <213> Saccharopolyspora erythraea

<400> 32
 ccgcaggagc gcgtgttcct ggaactcgct tgggaagcac ttgataacgc gggcatcgca 60
 ccgcacagcc tcagggacag ccggacgggc gtgttcttcg gagctatgtg gcacggctac 120
 gcgcagttcg cagccggagc cgtcgaccgc atcaccagc acaccgcgac cgggcacgac 180
 ctgagcatca tcccggccag gatcgctac ttctgggct tgcgcggccc ggacatgacc 240
 ctgaacaccg cgtgctcatc ggctttggtg gccatgcacc aggcacgcca aagcatcctg 300
 ctgggcgaat cctcggtcgc ctggtcggc gggatcagct tgttggtcgc gctggacagc 360
 atggtcgcca tgtcgcgggt cggagcgatg gcccggacg gccggtgcaa ggcattcgac 420
 tctcgcgcga acggctacgt gcgcggcgaa ggcggcggtg tcgtggtgct caaacggctg 480
 tcgcgcgctc tggccgatgg caaccgggtc tactgcgtcc tgcgcggcag cgcggtcaac 540
 aacgacggct tcagcaatgg ccttaccgcg ccgagcccgg cggcgaggga gcaggtactg 600
 cgcgacgcct acgccaacgc cggggtcgat ccggcacagg tcgactacgt cgagacccac 660
 gggaccggca c 671

<210> 33
 <211> 686
 <212> ADN
 <213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 33

```
ccgcaggagc gcgtgttcct cgagtcgtgc tgggaggcgc tggagcatgc tggatacga 60
actgcacgct accccggccg catcgggctg tgggccggcg cgggcttcaa cagctacctc 120
ctgaccaatc tcatgaacaa ccgcgccttt ttagagagcg tgggcatgta ccagatcttt 180
ctgagcaacg acaaggactt catcgccacc cgcacggctt acaagttaa cctgcgcggg 240
ccggcgatgg ccgtcggcac cgctgttcc acatcgctgg tggcggttca cgaagcttgc 300
caggcgctgc ggctgggca gtgtgacatg gactggccg gtgctgcgtc tgtcagcacg 360
ccctccggg agggctacct ctaccaggaa ggcattgatta tgagccgtga cggcgtctgc 420
cgcccgcttg acgccgacgc cgatggcacg gtgctgggca atggcggtgc ggtcgtggtg 480
ctcaagcggc tggacgaagc gctccgggac ggtgacacgg tctacgccgt gattcgtggc 540
acggcggtca acaacgacgg ctctgtcaag atcgggttca cggcgcccag cgccgagggg 600
cagagccggg tcgtgcggga cgccctgcgg gcggccgcgg tcccggcgga gagcgtgacc 660
tacgtcgaca cgcacggcac cggcac 686
```

<210> 34

<211> 689

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 34

```
ccccagcagc gcctgttcct cgagtgcgcg tgggaagcga tggagaacgc gggatatgcg 60
gcgcgaagct ataagggttc gatcggcgtt ttcgcgggat gcggcgtaa tacctacctg 120
ctgaacaacc tcgccaccgc ggagccgttc gatttctcac gcccctccgc gtaccagctg 180
ctgacggcca acgacaagga ttctctggcc acgcgtgtct cttacaagct gaacctccgc 240
gggccagcc tgacgggttca gacggcgtgc tccacctcgc tgggtgtcggg ggtgatggca 300
tgcgagagct tgcagcgcg cgctcggac attgccttgg ccgggggagt tgccatcaat 360
gttccgcagt ccgtggggta cctgcaccag ccgggcatga tcctgtcgcc cgacgggcgc 420
tgccgcgct tcgatgagtc cgctcaaggc acgggtgccg gcaacggcgc ggggtgtggtc 480
gtctcaagc gcttgagcgc cgctctggcc gatggcgaca cgatctacgc cgtcattcgc 540
ggagcggcta ttaataatga tggcgccgag cgcattgggt ttaccgctcc aggtgtggac 600
ggtcagacgc gattgattcg gcgcactcaa gagatggcgg gcgtgaagcc ggagtccatc 660
ggctacatgg acacccacgc caccggcac 689
```

<210> 35

<211> 671

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 35

```
ccgcagcagc gcctcttcct cgagggtggca tgggaagctt tggagcgtgc gggtcggccg 60
```

```

cccgacagtc tgcggggcag cgacaccgga gtgttcacgc ggatcagcac cgacgactac 120
agccggctga aacctaccga tccggcgctc attgacgcct ataccggtac cggaaccgcg 180
ttcagcactg ccgcccggacg gatctcctat ctgctggggt tgcagggacc gaacttcccc 240
gtcgacacgg cgtgctcttc ctcaactcgtg gcggttcacg tggcgtgccg cagcttgacg 300
tcgcgagagt gcagcatggc gctggccggc ggcgtgaacc tgattctggc gccggaaagc 360
acgatctact tctgccgcct gcgggccatg gcggccgatg gccgttgcaa aagtttcgct 420
gcctccggcg acggttacgg ccgcggcgag ggatgcggaa tgctgggtgct gaagcggctg 480
tccgatgcga cgcgtgacgg cgatcgtatt ctggcgctga ttccgcgatc ggccgtcaac 540
cacggcgccc gcagcaacgg cctcacggcg ccgaacggtc cggcgcgagg agccgtgatt 600
cgggcggcgc tcaagaacgc cggcatggcc ccgcgcgatg tcgattacgt ggacaccac 660
ggcaccggca c

```

671

<210> 36

<211> 758

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 36

```

ccgcaggagc gcgtcttcct cgaacgcatt gacggtttcg atgcggaatt cttcggcatc 60
tccccccgcg aagctctgaa catggatccg cagcagcggc tgctgctgga agtgtgctgg 120
gaagcggcag aggacgccgg catctctccc ggccctctgg cgggcagcgc gaccggcgctc 180
tttgccggct cctgcgcca ggacttcgga ctgtttcagt acgccgaccc tgcccgcac 240
ggagcttggc cgggttcggc cgtggcgcat agcatgttgg ccaatcgcat ctctatctg 300
ctcgacctgc gcggtccgag catggcggtc gatacggcct gctcctccgc gctcgtcgcc 360
gtccatctgg cttgccaaag cctgcgcccg cgcgaatgcg atgcggcatt cgcggcgga 420
gtgaacttga tctgactcc cgaggcatg atcgcttgt cgaaggctcg catgttggcg 480
cccgacggac gctgcaagac gttcgacgcc gcagccgacg gttatgtgcg cggcgagggc 540
tgccgcatcg tgctgctgaa gcggtctctc gatgcgctgg ccgatggcga tgccatctgt 600
gcagtcaccc gcggctcggc aatcaatcag gacggacgga gcaatggcat cacggcgccg 660
aatctgcagg cgcagaaggc ggtcctgcaa gaggcggtgg ccaacgcgca catcgatcca 720
tcccacgtat cgttgatcga cagcacggc accggcac

```

758

<210> 37

<211> 704

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 37

```

ccgcagcagc gcgtgttcct cgagtgcgcc tgggaggcgg tggaaagcgc gggctacgat 60
cccgaaaaat atcccggcct gatcggagtt ttgcgccggg ccagcatcaa cagctatttc 120
ctttataacc tcgcgcacaa ccgggaattc gtgcgccgca tggcggggga gtaccaagt 180
ggcgagtacc agacgatcct cggaaacgac aaggactacc tcccactcg cgtctcctac 240

```

```

aaattgaacc tgcgcggccc cagcctggcc gtgcagtcgg cctgctcgac cggcctcgtc 300
gccgtttgtc aggccattca aaatctgcag acttatcagt gcgatatggc cctcgcgggc 360
ggcatctcga ttctgtttcc gcaaaagcgc gactaccgct tcaccgacga aggaatggtc 420
tctcgcgacg gtcactggcg cccgttcgac gccagcgcgc aaggcacggt ctctggcaac 480
ggggccggcg tcgtcctgat gaaaagattg gccgacgcag tgaccgatcg ggacacgatc 540
ctcgccgtga ttagggggcg tgccgtgaac aacgacggcg gcgtcaaaat ggggttacacg 600
gcgcccagtg ccgaagggtca ggcgagggcc atcacctgg ccctcgcgct cgctggcgctc 660
agccccgaga ccatcacttg catggacacc cacggcaccg gcac 704

```

<210> 38

<211> 680

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 38

```

ccccagcagc gcgtgttcct cgaatgcgcc tgggcggcgc tggagcgccg ccggatatca 60
gggcgacacc ttccacgggtg tccatcggcg gtctatgcct caagcggctt taacacctat 120
cttctgaacc tgcattgcaa tgccgcgggtg cgccaatcga tcagcccgtt tgaactgttc 180
gtcgccaacg acaaggattt tctggcgacg cgcacggctt acaagctcaa tctgcgcggc 240
ccggccatga cagtgcagac ggctgctcc tcactgttgg ttgccgttca tctgcgcggc 300
caaagcctcc tagcggggcg atgcgatatt gcgctcgcgg gcggcatcac ggtttcccgt 360
tcgcatggat atgtggcgcg cgaagggtga atattgtctc ctgacgggca ttgccggggc 420
ttcgatgcgg atgccggcgg aaccgttcca ggcagcggcg tcggcgttgt cgtgctcaag 480
cgtctcgaag atgcgcttgc agacggcgat acgatcgacg ccgctcatcat cggttcggcc 540
atcaacaatg atggcgcgct gaaggcgagc tttaccgcac cgcaggtgga cagccaggcc 600
ttggtcatca gcgagggcca tgcagctgcc ggaatatcgg ccgattccat cggttatatg 660
gacacccacg gcaccgggac 680

```

<210> 39

<211> 671

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 39

```

ccgcagcagc gcctcttcct cgagctcacc tgggaagcgc tgggaagatgc cggcatcccg 60
ccgtccacga ttgccggcac gaatgtcggc gttttcatgg gcgcgtcgca ggctgactac 120
ggccacaagt tcttcagcga ccacgccgtc gcggattccc atttcgccac cggcacctcg 180
ctggcggtcg tcgccaatcg catttcctac atctacgacc tgcgcggccc aagcctcact 240
gtagacacgg cgtgctcgte gtcgctcgte gcgctgcatc aggcgggtgga agcgtccgc 300
tcggggcgga tcgaaacagc cattgtcggc ggcattaacg ttatcgccag cccggcgctc 360
ttcatcgctt tctcgcagge ctcgatgctg tcgccgacgg ggttgtgcca ggctttctcc 420
gccaaggccg atggctttgt ccgcggcgag ggcggcacgg ttttcgtcct gcgcaaggcg 480

```

```

gcgcacgccc atggcagccc caaccggtg cgcggggtca ttctcgccac cgacgtcaat 540
tccgacgggc gtaccaacgg catctcgctg ccatcgggcg aagcgcagga agtcctctg 600
caacgcgtct attcacgcgc atcgatcgat ccgaaccgcc tggctttcgt cgacaccac 660
gggaccggca c 671

```

<210> 40

<211> 764

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 40

```

ccgcagcagc gcgtgttcct cgacggcatc gaccggttcg atccgcgtca ctctcgcatc 60
acgcgcgcgc aggcgatcag catggacccg cagcagcggc tcctgctcga ggtcacgtgg 120
gaagcgctgg agcgcgccgg cgtggcgccc gatcgctga ccggatccga caccggcgtc 180
ttcatcgcca tcagaccaa cgactacggc cagatcctgc tgcgcgcctc ggaccagatc 240
gatccgggga tgtacttcgg caccggcaac ctggtgaacg cggcggcggg acgcctctcg 300
tacgtcctcg gcctgcaggg tccgagcatg gcggtcgaca ccgcattgcc gtcgtcgctg 360
gtggcgattc atctcgctg tccagagcctg cgcaaccgcg agtgccgcgt ggcgtcgcc 420
ggcggcgcca acctggtgct cgtcccggaa gtgacggtca actgctgccg cgccaagatg 480
ctcgcgcctg acgggcgcgt caagacgttc gacgcgcggc cggacggcta cgtccgcggc 540
gaagggcgcc cggtgatcgt gctgaagcgg ctctccgacg cgtggcgga cggcgatccg 600
atcgtcgcgc tgatccgcgg atccgcggtc aatcaggacg gccgcagcgg cggcttcacc 660
gcgcgaacg aactggcgca gcaggcgggtg atccggaccg cgtcgcgggc agcgggcgtc 720
gccgcgtccg acatcggtta cgtggacacg caccggaccg ggac 764

```

<210> 41

<211> 763

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 41

```

ccgcagcagc gcgtgttcct cgacggcatc gaccgcttcg atccgcagtt ttctcgggatc 60
gcgcgcgcgc aagcggccgg catcgatccg cagcagcggc tgctgctcga gacgacgtgg 120
gaagcgctgg aagacgccgg gacgtcgccg gaaaagctgc agggaaacccc ggccggcggtg 180
ttcgtcgcca tcaacagcat cgactacgcg acgctgcagc tgcagaactg cgatctggcc 240
agcatcgacg cctattcgct ctccggcagc gcgcacagca tcgcggccgg gcggtcgcc 300
tacgtgctcg gcctgcaggg gccggcgatg gcggtcgaca ccgcctgctc gtcgtcgctg 360
gtcgcgatcc acctggcggt ccagagcctg cgcaacgacg actgccgcgt cgccgtggcc 420
ggcggcggtg acgtcacgct gacgccgatc aacatggtcg tgttctcgaa gctgcgcatg 480
ctggcgcgcg acggcaagtg caagacgttc gacggccggc gcgacggatt cgtcgaaggc 540
gagggtcgcg cggtcacgtc cctcaagcgg ttgtcgacag cgcttgccga caaggatcgg 600
atcctcgcgc tgggtcgcgcg ttccggcggtc aaccaggacg gcgcgagcag cggcttcacc 660

```

gcgccgaacg gtccggcgca ggaagcggtc atccgcgcgg cgttgaagcg ggccggcggtg 720
 cagccggcgg aggtcggcta cgtggacacc caccgcaccg gca 763

<210> 42

<211> 668

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 42

ccgcaggagc gcgtgctgct ggaatcctcg tggcatgcmc tggaagacgc cggctatgcc 60
 ggcgaaagca tcgccggcgc gcgctgcggc gtgtacatgg gcttcaacgg cggcgactac 120
 ggcgacctgc tgtacggcca gccgtcgtcg ccgccgcacg cgatgtgggg caacgccgcc 180
 tcggtgctgt cggcgcgcat cgcctattac ctggacctgc aaggcccggc gatcaccctc 240
 gacaccgcct gttcgagctc gttggtcgcg gtgcatctgg cctgccaggg gctgtggacc 300
 ggcgagaccg atctggccct ggccggcggc gtgtggatcc agtgcacgcc cggattcctg 360
 atctctcca gccgcgccgg catgctctcg ccgaccggcc agtgccgcgc gttcggcgcc 420
 ggcgccgacg gcttcgtgcc gtccgaaggc gtcggcgctg tcgtgctcaa gcgcctgcag 480
 gacgcgctcg acgccggcga ccacatntac ggcgatgatcc gcggcagcgc gatcaaccag 540
 gacggcgcca gcaacggcat caccgcgcgg agcgccgccg cccaggagcg cttgcagcgc 600
 cacgtctacg acagcttcgg catcgacgcc tcgcgcctgc agatgatcga ggcccacggc 660
 accggcac 668

<210> 43

<211> 671

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 43

ccgcaggagc gcgtgctgct ggaggtgact tgggaggcac tcgaagacgc cggccaagac 60
 gtggaccgctc tggccgggcg gcccgtcggc gtcttcgctc ggatctcgtc gaacgattac 120
 ggccagcttc agaacggcga cccggccgac gtggacgcct acgtcggcac cggtaacgcg 180
 ctgagcatcg ccgccaaccg actcagctac acgtttgact ttcgcggccc gagtctggcg 240
 gtggacacgg cgtgctcgtc ttcactcgtc gcgatccatc tcgcctgcca gagcgttcgc 300
 cgcggtgaag cggaaactcg cgtgcgcggc ggcgtaact tgattctgac ccccggcctg 360
 acggtgaatt tcaccgcgc cggcatgatg gcgcctgacg gccggtgcaa gacgttcgac 420
 gcggccgcca acggctacgt gcgcggcgaa ggcccgggcg tcgtcgtgct caagccgctg 480
 gccagggcta tcgccgacgg cgaccggatc tacgcgatcg tccgtggcag cgccgtcaac 540
 caggacggcc gttccaacgg cctcaccgcc ccgaaccgac aggcccaaga ggtcgtgctg 600
 cgggcgcgct atcgtgacgc gggcatcagc ccggccgatg tcgacgccgt cgaggccac 660
 ggcaccggca c 671

<210> 44
 <211> 707
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence : organisme du sol

<400> 44
 ccccgagcgc gcgtgttcct cgaggacgcg actgaggtcg acgtggatgc gctttcagac 60
 ggcgaagacg tcgtgatcgc cggcatcatg cagcacatcg aggaggccgg catccactcg 120
 ggcgattcat cgtgcgtgct tccgccggtc gacatcccgc cgaaggcgct gcagacgac 180
 cgcgatcaca cgttcaagct cgcgcgcgcg ttgaaggta cgggcctgat gaacgtgcag 240
 tacgcgattc agcgcgacaa ggtctacgtg attgaggtaa accctagggc ttctcgaact 300
 gtcccgtatg tctcgaaggc gacaggcgtg ccgctggcga aggtcgcgtc acgcttgatg 360
 accggacgca aactgcacga gctgttgccg gaaggggtcg agcgcggctg gatcaccacc 420
 gcgggcgaga atttctacgt gaagtcgccg gtcttcccgt ggggtaagtt cccgggcgtt 480
 gacactgtgc tcgggccgga gatgaaatcg accggcgaag tcatgggcgt cgccgacaac 540
 ttccggcgagg ccttcgcca ggcacagatc gccgccggca catacctgcc gaccgaaggt 600
 accgtcttca tctcgggtcaa cgaccgtgac aaaggcaacg tcattcagct ggcgcagcgt 660
 ttctccgaac tcggtttcgg cattgtcgac acgcacggca ccgggac 707

<210> 45
 <211> 225
 <212> PRT
 <213> Streptomyces ambofaciens

<400> 45
 Pro Gln Gln His Val Phe Leu Glu Thr Val Trp Glu Thr Phe Glu Ser
 1 5 10 15
 Ala Gly Val Asp Pro Arg Ala Val Arg Gly Arg Ser Val Gly Met Phe
 20 25 30
 Val Gly Thr Asn Gly Gln Asp Tyr Pro Val Val Leu Ala Gly Ser Ala
 35 40 45
 Asp Glu Gly Leu Asp Ala His Ala Ala Thr Gly Asn Ala Ala Ala Val
 50 55 60
 Leu Ser Gly Arg Val Ser Tyr Ala Phe Gly Leu Glu Gly Pro Ala Val
 65 70 75 80
 Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Leu Ala
 85 90 95
 Ala Gln Ala Leu Arg Arg Gly Glu Cys Asp Leu Ala Leu Ala Gly Gly
 100 105 110

Val	Ser	Glu	Met	Ser	Thr	Glu	Ala	Ala	Phe	Thr	Glu	Phe	Ala	Arg	Gln
115				120				125							
Gly	Gly	Leu	Ala	Asp	Asp	Gly	Arg	Cys	Lys	Ala	Phe	Ser	Ala	Asp	Ala
130				135				140							
Asp	Gly	Thr	Gly	Trp	Gly	Glu	Gly	Val	Gly	Val	Leu	Leu	Val	Glu	Arg
145				150				155				160			
Leu	Ala	Asp	Ala	Arg	Arg	Asn	Gly	His	Arg	Ala	Leu	Ala	Leu	Val	Arg
				165				170				175			
Gly	Ser	Ala	Val	Asn	Gln	Asp	Gly	Ala	Ser	Asn	Gly	Leu	Thr	Ala	Pro
180				185				190							
Asn	Gly	Pro	Ser	Gln	Gln	Arg	Val	Ile	Arg	Gln	Ala	Leu	Ala	Asp	Ala
195				200				205							
Arg	Leu	Ser	Pro	Ser	Glu	Val	Asp	Ala	Val	Glu	Thr	His	Gly	Thr	Gly
210				215				220							
Thr															
225															

```
<210> 46
<211> 207
<212> PRT
<213> Streptomyces ambofaciens
```

```

<400> 46
Ala Ser Trp Glu Ala Val Glu Arg Ala Gly Ile Asp Met Arg Thr Leu
  1             5             10             15
Arg Gly Gly Arg Thr Gly Val Phe Ala Gly Val Met Tyr His Asp Tyr
          20             25             30
Pro Ser Val Val Asp Pro Glu Ala Leu Asp Gly Tyr Leu Gly Thr Ala
      35             40             45
Asn Ala Gly Ser Val Leu Ser Gly Arg Ile Ala Tyr Thr Phe Gly Leu
  50             55             60
Gln Gly Pro Ala Val Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val
  65             70             75             80
Ala Leu His Leu Ala Ala Gln Ala Leu Pro Ala Gly Glu Cys Glu Leu
          85             90             95

```

Ala Leu Val Gly Gly Val Thr Val Met Ser Gly Pro Met Met Phe Ala
 100 105 110

Gly Phe Gly Leu Glu Asp Gly Ser Ala Ala Asp Gly Arg Cys Lys Ala
 115 120 125

Phe Ala Ala Ala Ala Asp Gly Thr Gly Trp Gly Glu Gly Val Gly Val
 130 135 140

Leu Leu Val Glu Arg Leu Ser Asp Ala Arg Arg His Gly His Arg Val
 145 150 155 160

Leu Ala Val Val Arg Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Gly
 165 170 175

Gly Leu Thr Ala Pro Asn Gly Pro Ala Gln Gln Arg Val Ile Arg Gln
 180 185 190

Ala Leu Ala Ser Ala Ala Leu Val Pro Ala Glu Val Asp Ala Val
 195 200 205

<210> 47
 <211> 223
 <212> PRT
 <213> Saccharopolyspora erythraea

<400> 47
 Pro Gln Glu Arg Val Phe Leu Glu Leu Ala Trp Glu Ala Leu Asp Asn
 1 5 10 15

Ala Gly Ile Ala Pro His Ser Leu Arg Asp Ser Arg Thr Gly Val Phe
 20 25 30

Phe Gly Ala Met Trp His Gly Tyr Ala Gln Phe Ala Ala Gly Ala Val
 35 40 45

Asp Arg Ile Thr Gln His Thr Ala Thr Gly His Asp Leu Ser Ile Ile
 50 55 60

Pro Ala Arg Ile Ala Tyr Phe Leu Gly Leu Arg Gly Pro Asp Met Thr
 65 70 75 80

Leu Asn Thr Ala Cys Ser Ser Ala Leu Val Ala Met His Gln Ala Arg
 85 90 95

Gln Ser Ile Leu Leu Gly Glu Ser Ser Val Ala Leu Val Gly Gly Ile
 100 105 110

Ser Leu Leu Val Ala Leu Asp Ser Met Val Ala Met Ser Arg Phe Gly
 115 120 125
 Ala Met Ala Pro Asp Gly Arg Cys Lys Ala Phe Asp Ser Arg Ala Asn
 130 135 140
 Gly Tyr Val Arg Gly Glu Gly Gly Gly Val Val Val Leu Lys Pro Leu
 145 150 155 160
 Ser Arg Ala Leu Ala Asp Gly Asn Pro Val Tyr Cys Val Leu Arg Gly
 165 170 175
 Ser Ala Val Asn Asn Asp Gly Phe Ser Asn Gly Leu Thr Ala Pro Ser
 180 185 190
 Pro Ala Ala Gln Glu Gln Val Leu Arg Asp Ala Tyr Ala Asn Ala Gly
 195 200 205
 Val Asp Pro Ala Gln Val Asp Tyr Val Glu Thr His Gly Thr Gly
 210 215 220

<210> 48

<211> 211

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 48

Ser Cys Trp Glu Ala Leu Glu His Ala Gly Tyr Asp Thr Ala Arg Tyr
 1 5 10 15
 Pro Gly Arg Ile Gly Leu Trp Ala Gly Ala Gly Phe Asn Ser Tyr Leu
 20 25 30
 Leu Thr Asn Leu Met Asn Asn Arg Ala Phe Leu Glu Ser Val Gly Met
 35 40 45
 Tyr Gln Ile Phe Leu Ser Asn Asp Lys Asp Phe Ile Ala Thr Arg Thr
 50 55 60
 Ala Tyr Lys Leu Asn Leu Arg Gly Pro Ala Met Ala Val Gly Thr Ala
 65 70 75 80
 Cys Ser Thr Ser Leu Val Ala Val His Glu Ala Cys Gln Ala Leu Arg
 85 90 95

Leu Gly Glu Cys Asp Met Ala Leu Ala Gly Ala Ala Ser Val Ser Thr
 100 105 110
 Pro Leu Arg Glu Gly Tyr Leu Tyr Gln Glu Gly Met Ile Met Ser Arg
 115 120 125
 Asp Gly Val Cys Arg Pro Phe Asp Ala Asp Ala Asp Gly Thr Val Leu
 130 135 140
 Gly Asn Gly Val Ala Val Val Val Leu Lys Arg Leu Asp Glu Ala Leu
 145 150 155 160
 Arg Asp Gly Asp Thr Val Tyr Ala Val Ile Arg Gly Thr Ala Val Asn
 165 170 175
 Asn Asp Gly Ser Val Lys Ile Gly Phe Thr Ala Pro Ser Ala Glu Gly
 180 185 190
 Gln Ser Arg Val Val Arg Asp Ala Leu Arg Ala Ala Ala Val Pro Ala
 195 200 205
 Glu Ser Val
 210

<210> 49

<211> 229

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 49

Pro Gln Gln Arg Leu Phe Leu Glu Cys Ala Trp Glu Ala Met Glu Asn
 1 5 10 15
 Ala Gly Tyr Ala Ala Arg Ser Tyr Lys Gly Ser Ile Gly Val Phe Ala
 20 25 30
 Gly Cys Gly Val Asn Thr Tyr Leu Leu Asn Asn Leu Ala Thr Ala Glu
 35 40 45
 Pro Phe Asp Phe Ser Arg Pro Ser Ala Tyr Gln Leu Leu Thr Ala Asn
 50 55 60
 Asp Lys Asp Phe Leu Ala Thr Arg Val Ser Tyr Lys Leu Asn Leu Arg
 65 70 75 80

Gly	Pro	Ser	Leu	Thr	Val	Gln	Thr	Ala	Cys	Ser	Thr	Ser	Leu	Val	Ser	
				85					90					95		
Val	Val	Met	Ala	Cys	Glu	Ser	Leu	Gln	Arg	Gly	Ala	Ser	Asp	Ile	Ala	
				100					105					110		
Leu	Ala	Gly	Gly	Val	Ala	Ile	Asn	Val	Pro	Gln	Ser	Val	Gly	Tyr	Leu	
				115					120					125		
His	Gln	Pro	Gly	Met	Ile	Leu	Ser	Pro	Asp	Gly	Arg	Cys	Arg	Ala	Phe	
				130					135					140		
Asp	Glu	Ser	Ala	Gln	Gly	Thr	Val	Pro	Gly	Asn	Gly	Ala	Gly	Val	Val	
				145					150					155		
Val	Leu	Lys	Arg	Leu	Ser	Arg	Ala	Leu	Ala	Asp	Gly	Asp	Thr	Ile	Tyr	
				165					170					175		
Ala	Val	Ile	Arg	Gly	Ala	Ala	Ile	Asn	Asn	Asp	Gly	Ala	Glu	Arg	Met	
				180					185					190		
Gly	Phe	Thr	Ala	Pro	Gly	Val	Asp	Gly	Gln	Thr	Arg	Leu	Ile	Arg	Arg	
				195					200					205		
Thr	Gln	Glu	Met	Ala	Gly	Val	Lys	Pro	Glu	Ser	Ile	Gly	Tyr	Met	Asp	
				210					215					220		
Thr	His	Gly	Thr	Gly												
225																

<210> 50

<211> 223

<212> PRT

<213> Organime. Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 50

Pro Gln Gln Arg Leu Phe Leu Glu Val Ala Trp Glu Ala Leu Glu Arg
1 5 10 15

Ala Gly Arg Pro Pro Asp Ser Leu Ala Gly Ser Asp Thr Gly Val Phe
20 25 30

Ile Gly Ile Ser Thr Asp Asp Tyr Ser Arg Leu Lys Pro Thr Asp Pro
35 40 45

Ala Leu Ile Asp Ala Tyr Thr Gly Thr Gly Thr Ala Phe Ser Thr Ala
50 55 60

Ala Gly Arg Ile Ser Tyr Leu Leu Gly Leu Gln Gly Pro Asn Phe Pro
65 70 75 80

Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Leu Ala Cys
85 90 95

Arg Ser Leu Gln Ser Arg Glu Cys Ser Met Ala Leu Ala Gly Gly Val
100 105 110

Asn Leu Ile Leu Ala Pro Glu Ser Thr Ile Tyr Phe Cys Arg Leu Arg
115 120 125

Ala Met Ala Ala Asp Gly Arg Cys Lys Ser Phe Ala Ala Ser Ala Asp
130 135 140

Gly Tyr Gly Arg Gly Glu Gly Cys Gly Met Leu Val Leu Lys Arg Leu
145 150 155 160

Ser Asp Ala Thr Arg Asp Gly Asp Arg Ile Leu Ala Leu Ile Arg Gly
165 170 175

Ser Ala Val Asn His Gly Gly Arg Ser Asn Gly Leu Thr Ala Pro Asn
180 185 190

Gly Pro Ala Gln Glu Ala Val Ile Arg Ala Ala Leu Lys Asn Ala Gly
195 200 205

Met Ala Pro Ala Asp Val Asp Tyr Val Asp Thr His Gly Thr Gly
210 215 220

<210> 51

<211> 252

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 51

Pro Gln Glu Arg Val Phe Leu Glu Arg Ile Asp Gly Phe Asp Ala Glu
1 5 10 15

Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Asn Met Asp Pro Gln Gln
20 25 30

Arg Leu Leu Leu Glu Val Cys Trp Glu Ala Ala Glu Asp Ala Gly Ile
 35 40 45
 Ser Pro Gly Pro Leu Ala Gly Ser Ala Thr Gly Val Phe Ala Gly Ser
 50 55 60
 Cys Ala Gln Asp Phe Gly Leu Phe Gln Tyr Ala Asp Pro Ala Arg Ile
 65 70 75 80
 Gly Ala Trp Ser Gly Ser Gly Val Ala His Ser Met Leu Ala Asn Arg
 85 90 95
 Ile Ser Tyr Leu Leu Asp Leu Arg Gly Pro Ser Met Ala Val Asp Thr
 100 105 110
 Ala Cys Ser Ser Ala Leu Val Ala Val His Leu Ala Cys Gln Ser Leu
 115 120 125
 Arg Arg Arg Glu Cys Asp Ala Ala Phe Ala Gly Gly Val Asn Leu Ile
 130 135 140
 Leu Thr Pro Glu Gly Met Ile Ala Leu Ser Lys Ala Arg Met Leu Ala
 145 150 155 160
 Pro Asp Gly Arg Cys Lys Thr Phe Asp Ala Ala Ala Asp Gly Tyr Val
 165 170 175
 Arg Gly Glu Gly Cys Gly Ile Val Leu Leu Lys Arg Leu Ser Asp Ala
 180 185 190
 Leu Ala Asp Gly Asp Ala Ile Cys Ala Val Ile Arg Gly Ser Ala Ile
 195 200 205
 Asn Gln Asp Gly Arg Ser Asn Gly Ile Thr Ala Pro Asn Leu Gln Ala
 210 215 220
 Gln Lys Ala Val Leu Gln Glu Ala Val Ala Asn Ala His Ile Asp Pro
 225 230 235 240
 Ser His Val Ser Leu Ile Asp Thr His Gly Thr Gly
 245 250

<210> 52

<211> 234

<212> PRT

<213> Organime Inconnu

<220>

<223> Origine de la séquence : organisme du sol

<400> 52

Pro	Gln	Gln	Arg	Val	Phe	Leu	Glu	Cys	Ala	Trp	Glu	Ala	Val	Glu	Ser	1	5	10	15
Ala	Gly	Tyr	Asp	Pro	Glu	Lys	Tyr	Pro	Gly	Leu	Ile	Gly	Val	Phe	Ala	20	25	30	
Gly	Ala	Ser	Ile	Asn	Ser	Tyr	Phe	Leu	Tyr	Asn	Leu	Ala	His	Asn	Arg	35	40	45	
Glu	Phe	Val	Ala	Arg	Met	Ala	Gly	Glu	Tyr	Gln	Val	Gly	Glu	Tyr	Gln	50	55	60	
Thr	Ile	Leu	Gly	Asn	Asp	Lys	Asp	Tyr	Leu	Pro	Thr	Arg	Val	Ser	Tyr	65	70	75	80
Lys	Leu	Asn	Leu	Arg	Gly	Pro	Ser	Leu	Ala	Val	Gln	Ser	Ala	Cys	Ser	85	90	95	
Thr	Gly	Leu	Val	Ala	Val	Cys	Gln	Ala	Ile	Gln	Asn	Leu	Gln	Thr	Tyr	100	105	110	
Gln	Cys	Asp	Met	Ala	Leu	Ala	Gly	Gly	Ile	Ser	Ile	Ser	Phe	Pro	Gln	115	120	125	
Lys	Arg	Asp	Tyr	Arg	Phe	Thr	Asp	Glu	Gly	Met	Val	Ser	Arg	Asp	Gly	130	135	140	
His	Cys	Arg	Pro	Phe	Asp	Ala	Ser	Ala	Gln	Gly	Thr	Val	Phe	Gly	Asn	145	150	155	160
Gly	Ala	Gly	Val	Val	Leu	Met	Lys	Arg	Leu	Ala	Asp	Ala	Val	Thr	Asp	165	170	175	
Arg	Asp	Thr	Ile	Leu	Ala	Val	Ile	Arg	Gly	Ala	Ala	Val	Asn	Asn	Asp	180	185	190	
Gly	Gly	Val	Lys	Met	Gly	Tyr	Thr	Ala	Pro	Ser	Ala	Glu	Gly	Gln	Ala	195	200	205	
Glu	Ala	Ile	Thr	Leu	Ala	Leu	Ala	Leu	Ala	Gly	Val	Ser	Pro	Glu	Thr	210	215	220	
Ile	Thr	Cys	Met	Asp	Thr	His	Gly	Thr	Gly	225	230								

<210> 53
 <211> 226
 <212> PRT
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :organisme du sol

<400> 53

Pro	Gln	Gln	Arg	Val	Phe	Leu	Glu	Cys	Ala	Trp	Ala	Ala	Leu	Glu	Arg	1	5	10	15
Arg	Arg	Ile	Ser	Gly	Arg	His	Leu	Pro	Arg	Cys	Pro	Ser	Ala	Val	Tyr	20	25	30	
Ala	Ser	Ser	Gly	Phe	Asn	Thr	Tyr	Leu	Leu	Asn	Leu	His	Ala	Asn	Ala	35	40	45	
Ala	Val	Arg	Gln	Ser	Ile	Ser	Pro	Phe	Glu	Leu	Phe	Val	Ala	Asn	Asp	50	55	60	
Lys	Asp	Phe	Leu	Ala	Thr	Arg	Thr	Ala	Tyr	Lys	Leu	Asn	Leu	Arg	Gly	65	70	75	80
Pro	Ala	Met	Thr	Val	Gln	Thr	Ala	Cys	Ser	Ser	Ser	Leu	Val	Ala	Val	85	90	95	
His	Val	Ala	Ala	Gln	Ser	Leu	Leu	Ala	Gly	Glu	Cys	Asp	Ile	Ala	Leu	100	105	110	
Ala	Gly	Gly	Ile	Thr	Val	Ser	Arg	Ser	His	Gly	Tyr	Val	Ala	Arg	Glu	115	120	125	
Gly	Gly	Ile	Leu	Ser	Pro	Asp	Gly	His	Cys	Arg	Ala	Phe	Asp	Ala	Asp	130	135	140	
Ala	Gly	Gly	Thr	Val	Pro	Gly	Ser	Gly	Val	Gly	Val	Val	Val	Leu	Lys	145	150	155	160
Arg	Leu	Glu	Asp	Ala	Leu	Ala	Asp	Gly	Asp	Thr	Ile	Asp	Ala	Val	Ile	165	170	175	
Ile	Gly	Ser	Ala	Ile	Asn	Asn	Asp	Gly	Ala	Leu	Lys	Ala	Ser	Phe	Thr	180	185	190	
Ala	Pro	Gln	Val	Asp	Ser	Gln	Ala	Leu	Val	Ile	Ser	Glu	Ala	His	Ala	195	200	205	
Ala	Ala	Gly	Ile	Ser	Ala	Asp	Ser	Ile	Gly	Tyr	Met	Asp	Thr	His	Gly	210	215	220	

Thr Gly
225

<210> 54

<211> 223

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence : organisme du sol

<400> 54

Pro Gln Gln Arg Leu Phe Leu Glu Leu Thr Trp Glu Ala Leu Glu Asp
1 5 10 15

Ala Gly Ile Pro Pro Ser Thr Ile Ala Gly Thr Asn Val Gly Val Phe
20 25 30

Met Gly Ala Ser Gln Ala Asp Tyr Gly His Lys Phe Phe Ser Asp His
35 40 45

Ala Val Ala Asp Ser His Phe Ala Thr Gly Thr Ser Leu Ala Val Val
50 55 60

Ala Asn Arg Ile Ser Tyr Ile Tyr Asp Leu Arg Gly Pro Ser Leu Thr
65 70 75 80

Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Gln Ala Val
85 90 95

Glu Ala Leu Arg Ser Gly Arg Ile Glu Thr Ala Ile Val Gly Gly Ile
100 105 110

Asn Val Ile Ala Ser Pro Ala Ser Phe Ile Ala Phe Ser Gln Ala Ser
115 120 125

Met Leu Ser Pro Thr Gly Leu Cys Gln Ala Phe Ser Ala Lys Ala Asp
130 135 140

Gly Phe Val Arg Gly Glu Gly Gly Thr Val Phe Val Leu Arg Lys Ala
145 150 155 160

Ala His Ala His Gly Ser Arg Asn Pro Val Arg Gly Leu Ile Leu Ala
165 170 175

Thr Asp Val Asn Ser Asp Gly Arg Thr Asn Gly Ile Ser Leu Pro Ser
180 185 190

Ala Glu Ala Gln Glu Val Leu Leu Gln Arg Val Tyr Ser Arg Ala Ser
 195 200 205

Ile Asp Pro Asn Arg Leu Ala Phe Val Asp Thr His Gly Thr Gly
 210 215 220

<210> 55

<211> 254

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence : organisme du sol

<400> 55

Pro Gln Gln Arg Val Phe Leu Asp Gly Ile Asp Arg Phe Asp Pro Arg
 1 5 10 15

His Phe Ala Ile Thr Pro Arg Glu Ala Ile Ser Met Asp Pro Gln Gln
 20 25 30

Arg Leu Leu Leu Glu Val Thr Trp Glu Ala Leu Glu Arg Ala Gly Val
 35 40 45

Ala Pro Asp Arg Leu Thr Gly Ser Asp Thr Gly Val Phe Ile Gly Ile
 50 55 60

Ser Thr Asn Asp Tyr Gly Gln Ile Leu Leu Arg Ala Ser Asp Gln Ile
 65 70 75 80

Asp Pro Gly Met Tyr Phe Gly Thr Gly Asn Leu Leu Asn Ala Ala Ala
 85 90 95

Gly Arg Leu Ser Tyr Val Leu Gly Leu Gln Gly Pro Ser Met Ala Val
 100 105 110

Asp Thr Ala Cys Pro Ser Ser Leu Val Ala Ile His Leu Ala Cys Gln
 115 120 125

Ser Leu Arg Asn Arg Glu Cys Arg Met Ala Leu Ala Gly Gly Ala Asn
 130 135 140

Leu Val Leu Val Pro Glu Val Thr Val Asn Cys Cys Arg Ala Lys Met
 145 150 155 160

Leu Ala Pro Asp Gly Arg Cys Lys Thr Phe Asp Ala Ala Ala Asp Gly
 165 170 175

Tyr Val Arg Gly Glu Gly Ala Ala Val Ile Val Leu Lys Arg Leu Ser
 180 185 190
 Asp Ala Leu Ala Asp Gly Asp Pro Ile Val Ala Leu Ile Arg Gly Ser
 195 200 205
 Ala Val Asn Gln Asp Gly Arg Ser Gly Gly Phe Thr Ala Pro Asn Glu
 210 215 220
 Leu Ala Gln Gln Ala Val Ile Arg Thr Ala Leu Ala Ala Ala Gly Val
 225 230 235 240
 Ala Ala Ser Asp Ile Gly Tyr Val Asp Thr His Gly Thr Gly
 245 250

<210> 56

<211> 254

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 56

Pro Gln Gln Arg Val Phe Leu Asp Gly Ile Asp Arg Phe Asp Pro Gln
 1 5 10 15
 Phe Phe Gly Ile Ala Pro Arg Glu Ala Ala Gly Ile Asp Pro Gln Gln
 20 25 30
 Arg Leu Leu Leu Glu Thr Thr Trp Glu Ala Leu Glu Asp Ala Gly Thr
 35 40 45
 Ser Pro Glu Lys Leu Gln Gly Thr Pro Ala Gly Val Phe Val Gly Ile
 50 55 60
 Asn Ser Ile Asp Tyr Ala Thr Leu Gln Leu Gln Asn Cys Asp Leu Ala
 65 70 75 80
 Ser Ile Asp Ala Tyr Ser Leu Ser Gly Ser Ala His Ser Ile Ala Ala
 85 90 95
 Gly Arg Leu Ala Tyr Val Leu Gly Leu Gln Gly Pro Ala Met Ala Val
 100 105 110
 Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu Ala Cys Gln
 115 120 125

Ser Leu Arg Asn Asp Asp Cys Arg Val Ala Val Ala Gly Gly Val His
 130 135 140
 Val Thr Leu Thr Pro Ile Asn Met Val Val Phe Ser Lys Leu Arg Met
 145 150 155 160
 Leu Ala Ala Asp Gly Lys Cys Lys Thr Phe Asp Gly Arg Gly Asp Gly
 165 170 175
 Phe Val Glu Gly Glu Gly Cys Ala Val Ile Val Leu Lys Arg Leu Ser
 180 185 190
 His Ala Leu Ala Asp Lys Asp Arg Ile Leu Ala Leu Val Arg Gly Ser
 195 200 205
 Ala Val Asn Gln Asp Gly Ala Ser Ser Gly Leu Thr Ala Pro Asn Gly
 210 215 220
 Pro Ala Gln Glu Ala Val Ile Arg Ala Ala Leu Lys Arg Ala Gly Val
 225 230 235 240
 Gln Pro Ala Glu Val Gly Tyr Val Asp Thr His Gly Thr Gly
 245 250

<210> 57

<211> 222

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :organisme du sol

<400> 57

Pro Gln Glu Arg Val Leu Leu Glu Ser Ser Trp His Ala Leu Glu Asp
 1 5 10 15
 Ala Gly Tyr Ala Gly Glu Ser Ile Ala Gly Ala Arg Cys Gly Val Tyr
 20 25 30
 Met Gly Phe Asn Gly Gly Asp Tyr Gly Asp Leu Leu Tyr Gly Gln Pro
 35 40 45
 Ser Leu Pro Pro His Ala Met Trp Gly Asn Ala Ala Ser Val Leu Ser
 50 55 60
 Ala Arg Ile Ala Tyr Tyr Leu Asp Leu Gln Gly Pro Ala Ile Thr Leu
 65 70 75 80

Asp	Thr	Ala	Cys	Ser	Ser	Leu	Val	Ala	Val	His	Leu	Ala	Cys	Gln	85	90	95	
Gly	Leu	Trp	Thr	Gly	Glu	Thr	Asp	Leu	Ala	Leu	Ala	Gly	Gly	Val	Trp	100	105	110
Ile	Gln	Cys	Thr	Pro	Gly	Phe	Leu	Ile	Ser	Ser	Ser	Arg	Ala	Gly	Met	115	120	125
Leu	Ser	Pro	Thr	Gly	Gln	Cys	Arg	Ala	Phe	Gly	Ala	Gly	Ala	Asp	Gly	130	135	140
Phe	Val	Pro	Ser	Glu	Gly	Val	Gly	Val	Val	Val	Leu	Lys	Arg	Leu	Gln	145	150	155
Asp	Ala	Leu	Asp	Ala	Gly	Asp	His	Xaa	Tyr	Gly	Val	Ile	Arg	Gly	Ser	165	170	175
Ala	Ile	Asn	Gln	Asp	Gly	Ala	Ser	Asn	Gly	Ile	Thr	Ala	Pro	Ser	Ala	180	185	190
Ala	Ala	Gln	Glu	Arg	Leu	Gln	Arg	His	Val	Tyr	Asp	Ser	Phe	Gly	Ile	195	200	205
Asp	Ala	Ser	Arg	Leu	Gln	Met	Ile	Glu	Ala	His	Gly	Thr	Gly			210	215	220

```
<210> 58
<211> 223
<212> PRT
<213> Organisme Inconnu
```

<220>
<223> Origine de la séquence :organisme du sol

```

<400> 58
Pro Gln Glu Arg Val Leu Leu Glu Val Thr Trp Glu Ala Leu Glu Asp
 1                    5                        10                15
Ala Gly Gln Asp Val Asp Arg Leu Ala Gly Arg Pro Val Gly Val Phe
                20                    25                30
Val Gly Ile Ser Ser Asn Asp Tyr Gly Gln Leu Gln Asn Gly Asp Pro
    35                        40                45
Ala Asp Val Asp Ala Tyr Val Gly Thr Gly Asn Ala Leu Ser Ile Ala
    50                    55                60

```

Ala Asn Arg Leu Ser Tyr Thr Phe Asp Phe Arg Gly Pro Ser Leu Ala
 65 70 75 80
 Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu Ala Cys
 85 90 95
 Gln Ser Val Arg Arg Gly Glu Ala Glu Leu Ala Val Ala Ala Gly Val
 100 105 110
 Asn Leu Ile Leu Thr Pro Gly Leu Thr Val Asn Phe Thr Arg Ala Gly
 115 120 125
 Met Met Ala Pro Asp Gly Arg Cys Lys Thr Phe Asp Ala Ala Ala Asn
 130 135 140
 Gly Tyr Val Arg Gly Glu Gly Ala Gly Val Val Val Leu Lys Pro Leu
 145 150 155 160
 Ala Gln Ala Ile Ala Asp Gly Asp Pro Ile Tyr Ala Ile Val Arg Gly
 165 170 175
 Ser Ala Val Asn Gln Asp Gly Arg Ser Asn Gly Leu Thr Ala Pro Asn
 180 185 190
 Arg Gln Ala Gln Glu Val Val Leu Arg Ala Ala Tyr Arg Asp Ala Gly
 195 200 205
 Ile Ser Pro Ala Asp Val Asp Ala Val Glu Ala His Gly Thr Gly
 210 215 220

<210> 59

<211> 235

<212> PRT

<213> Organisme Inconnu

<220>

<223> Origine de la séquence : organisme du sol

<400> 59

Pro Gln Gln Arg Val Phe Leu Glu Asp Ala Thr Glu Val Asp Val Asp
 1 5 10 15

Ala Leu Ser Asp Gly Glu Asp Val Val Ile Ala Gly Ile Met Gln His
 20 25 30

Ile Glu Glu Ala Gly Ile His Ser Gly Asp Ser Ser Cys Val Leu Pro
 35 40 45

Pro Val Asp Ile Pro Pro Lys Ala Leu Gln Thr Ile Arg Asp His Thr
 50 55 60
 Phe Lys Leu Ala Arg Ala Leu Lys Val Ile Gly Leu Met Asn Val Gln
 65 70 75 80
 Tyr Ala Ile Gln Arg Asp Lys Val Tyr Val Ile Glu Val Asn Pro Arg
 85 90 95
 Ala Ser Arg Thr Val Pro Tyr Val Ser Lys Ala Thr Gly Val Pro Leu
 100 105 110
 Ala Lys Val Ala Ser Arg Leu Met Thr Gly Arg Lys Leu His Glu Leu
 115 120 125
 Leu Pro Glu Gly Val Glu Arg Gly Trp Ile Thr Thr Ala Gly Glu Asn
 130 135 140
 Phe Tyr Val Lys Ser Pro Val Phe Pro Trp Gly Lys Phe Pro Gly Val
 145 150 155 160
 Asp Thr Val Leu Gly Pro Glu Met Lys Ser Thr Gly Glu Val Met Gly
 165 170 175
 Val Ala Asp Asn Phe Gly Glu Ala Phe Ala Lys Ala Gln Ile Ala Ala
 180 185 190
 Gly Thr Tyr Leu Pro Thr Glu Gly Thr Val Phe Ile Ser Val Asn Asp
 195 200 205
 Arg Asp Lys Gly Asn Val Ile Gln Leu Ala Gln Arg Phe Ser Glu Leu
 210 215 220
 Gly Phe Gly Ile Val Asp Thr His Gly Thr Gly
 225 230 235

<210> 60

<211> 1269

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 60

taacaggaag aagcttgctt ctttgctgac gagtggcgga cgggtgagta acacgtggga 60
 acctgcctta tggttcgga taacgtctgg aaacggacgc taacaccgga tgtgcccttc 120

```

gggggaaagt ttacgccatg agagggggccc gcgtccgatt aggtagttgg tggggtaatg 180
gcccaccaag ccgacgatcg gtagctgggtc tgagaggatg atcagccaca ctgggactga 240
gacacggccc agactcctac gggaggcagc agtggggaat attggacaat gggggcaacc 300
ctgatccagc aatgccgcgt gagtgatgaa ggccttaggg ttgtaaagct ctttcgcacg 360
cgacgatgat gacggtagcg tgagaagaag ccccggttaa cttcgtgcca gcagccgcgg 420
taatacgaag ggggcgagcg ttgttcggaa ttaactgggcg taaagggcg gtaggcgggc 480
cgatcagtca gatgtgaaag ccccgggctc aacctgggaa ctgcatttga tactgtcggg 540
cttgagttcc ggagaggatg gtggaattcc cagtgtagag gtgaaattcg tagatattgg 600
gaagaacacc ggtggcgaag gcggccatct ggacggacac tgacgctgag gcgcgaaagc 660
gtggggagca aacaggatta gataccctgg tagtcacgc cgtaaacgat gaatgctaga 720
cgctgggggtg catgcacttc ggtgtcgccg ctaacgcatt aagcattccg cctggggagt 780
acggccgcaa gggttaaaact caaaggaatt gacggggggc cgcacaagcg gtggagcatg 840
tggtttaatt cgaagcaacg cgcagaacct taccaacct tgacatgtcc attgccggtc 900
cgagagattg gaccttcagt tcggctggat ggaacacagg tgetgcatgg ctgtcgtcag 960
ctcgtgtcgt gagatgttgg gttaagtccc gcaacgagcg caaccctac cgccagttgc 1020
catcattcag ttgggcactc tgggtggaact gccggtgaca agccggagga agggggggat 1080
gacgtcaagt cctcatggcc cttatgggtt gggctacaca cgtgctacaa tagcggtgac 1140
agtgggacgc gaagtgcgaa gatggagcaa atccccaaaa gccgtctcag ttcggattgc 1200
actctgcaac tcgggtgcat gaagttggaa tcgctagtaa tcgcggatca gcacgccgcg 1260
gtgaatacg                                     1269

```

<210> 61

<211> 1500

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 61

```

ttttaaaccg acggccagtg aattgtaata cgactcacta tagggcgaat tgggccctct 60
agatgcatgc tcgagcggcc gccagtgtga tggatatctg cagaattcgc ccttcaggcc 120
taacacatgc aagtgcgaac agggcttcgg ccctagtggc gcacgggtga gtaacacgtg 180
ggaacctgcc ttatggttcg ggataacgtc tggaaacgga cgtaaacacc ggatgtgccc 240
ttcgggggaa agtttacgcc atgagagggg cccgcgtccg attaggtagt tgggtgggta 300
atggcccacc aagccgacga tcggtagctg gtctgagagg atgatcagcc aactggggac 360
tgagacacgg ccagactcc tacgggaggc agcagtgggg aatattggac aatgggggca 420
acctgatcc agcaatgccg cgtgagtgat gaaggcctta gggttgtaaa gctctttcgc 480
acgcgacgat gatgacggta gcgtgagaag aagccccggc taacttcgtg ccagcagccg 540
cggtataacg aagggggcga gcgttggtcg gaattactgg gcgtaaaggg cgcgtaggcg 600
gcccgatcag tcagatgtga aagccccggg ctcaacctgg gaactgcatt tgatactgtc 660
gggcttgagt tccggagagg atggtggaat tccagtgta gaggtgaaat tcgtagatat 720
tgggaagaac accggtggcg aaggcgcca tctggacgga cactgacgct gaggcgcgaa 780
agcgtgggga gcaaacagga ttagataccc tggtagtcca cgccgtaaac gatgaatgct 840
agacgtggg gtgcatgcac ttcggtgtcg ccgctaacgc attaagcatt ccgcctgggg 900
agtacggccg caagggtaaa actcaaagga attgacgggg gcccgcaaa gcggtggagc 960
atgtggttta attcgaagca acgcgcagaa ccttaccac ccttgacatg tccattgccg 1020
gtccgagaga ttggaccttc agttcggctg gatggaacac aggtgctgca tggctgtcgt 1080
cagctcgtgt cgtgagatgt tgggttaagt cccgcaacga gcgcaacccc taccgccagt 1140

```

```

tgccatcatt cagttgggca ctctggtgga actgccggtg acaagccgga ggaaggcggg 1200
gatgacgtca agtcctcatg gcccttatgg gttgggctac acacgtgcta caatggcggg 1260
gacagtggga cgcgaagtcg caagatggag caaatcccca aaagccgtct cagttcggat 1320
tgcaactctgc aactcgggtg catgaagttg gaatcgctag taatcgcgga tcagcacgcc 1380
gcggtgaata cgttcccggg ccttgtagac accgcccag ggcgaattcc agcactgg 1440
cgcccggttac tagtggatcc gagctcggtg ccaagcttgg cgtaatcatg gtcatagctg 1500

```

<210> 62

<211> 1366

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 62

```

acgacggcca gtgaattgta atacgactca ctatagggcg aattgggccc tctagatgca 60
tgctcgagcg gccgccagtg tgatggatat ctgcagaatt cgcccttcag gcctaacaca 120
tgcaagtcga acgaaggctt cggccttagt ggcgcacggg tgagtaacac gtgggaacct 180
gcctttcggg tcggaataac gtctggaaac ggacgctaac accggatacg cccttcgggg 240
gaaagtccac gccgagagag gggcccgctg cggattaggt agttggtgag gtaatggctc 300
accaagcctt cgatccgtag ctggtctgag aggatgatca gccacactgg gactgagaca 360
cggcccagac tcctacggga ggcagcagtg gggaaatttg gacaatgggc gcaagcctga 420
tccagcaatg ccgcgtgagt gatgaaggcc ttagggttgt aaagctcttt cgcacgcgac 480
gatgatgacg gtagcgtgag aagaagcccc ggctaacttc gtgccagcag ccgcggtaat 540
acgaaggggg ctagcgttgt tcggaattac tgggcgtaaa gggcgcgtag gcggcctgct 600
tagtcagaag tgaaagcccc gggctcaacc tgggaatagc ttttgatact ggcaggcttg 660
agttccggag aggatggtgg aattcccagt gtagaggtag aattcgtaga tattgggaag 720
aacaccggtg gcgaaggcgg ccatctggac ggacactgac gctgaggcgc gaaagcgtgg 780
ggagcaaaca ggattagata ccctggtagt ccacgccgta aacgatgaat gctagacgtc 840
ggggtgcatg cacttcgggtg tcgccgctaa cgcattaagc attccgcctg gggagtacgg 900
ccgcaaggtt aaaactcaaa ggaattgacg ggggcccgca caagcgggtg agcatgtggt 960
ttaattcgaa gcaacgcgca gaaccttacc aacccttgac atgtccatta tgggcttcag 1020
agatgaggtc cttcagttcg gctgggtgga acacaggtgc tgcattggctg tcgtcagctc 1080
gtgtcgtgag atgttggtgt aagtcccgca acgagcgcaa cccctaccgt cagttgccat 1140
cattcagttg ggcactctgg tggaaaccgc ggtgacaagc cggaggaagg cggggatgac 1200
gtcaagtcct catggccctt atgggttggg ctacacacgt gctacaatgg cggtgacagt 1260
gggaagcgaa gtcgcgagat ggagcaaacc cccaaaagcc gtctcagttc ggatcgact 1320
ctgcaactcg agtgcgtgaa gttggaatcg ctagtaatcg cggatc 1366

```

<210> 63

<211> 1360

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 63

```

acagctatga ccatgattac gccaaagcttg gtaccgagct cggatccact agtaacggcc 60
gccagtgtgc tggaaattcgc ccttcaggcc taacacatgc aagtcgaacg ccccgcaagg 120
ggagtggcag acgggtgagt aacgcgtggg aacataccct ttcctgcgga atagctccgg 180
gaaactggaa ttaataccgc atacgcccta cgggggaaag atttatcggg gaaggattgg 240
cccgcgttgg attagctagt tgggtgggta aaggcctacc aaggcgacga tccatagctg 300
gtctgagagg atgatcagcc acattgggac tgagacacgg cccaaactcc tacgggaggc 360
agcagtgggg aatattggac aatgggcgca agcctgatcc agccatgccg cgtgagtgat 420
gaaggcctta gggttgtaaa gctctttcac cggagaagat aatgacggtg tccggagaag 480
aagccccggc taacttcgtg ccagcagccg cggtaatacg aagggggcta gtgttggtcg 540
gaattactgg gcgtaaagcg cacgtaggcg gatatttaag tcaggggtga aatcccagag 600
ctcaactctg gaactgcctt tgatactggg tatcttgagt atggaagagg taagtggaa 660
tccgagtgtg gaggtgaaat tcgtagatat tcggaggaa accagtggcg aaggcggtt 720
actggtccat tactgacgct gaggtgcgaa agcgtgggga gcaaacagga ttagataccc 780
tggtagtcca cgccgtaaac gatgaatgtt agccgtcggg cagtatactg ttcggtggcg 840
cagctaacgc attaaacatt ccgcctgggg agtacggtcg caagattaaa actcaaagga 900
attgacgggg gcccgcaaaa gcggtggagc atgtggttta attcgaagca acgcgcagaa 960
ccttaccagc tcttgacatt cggggtttgg gcagtggaga cattgtcctt cagttaggct 1020
ggccccagaa cagggtgctgc atggctgtcg tcagctcgcg tcgtgagatg ttgggttaag 1080
tcccgaacg agcgcaaccc tcgccttag ttgccagcat ttagttgggc actctaagg 1140
gactgccggg gataagccga gaggaagggt gggacgacgt caagtcctca tggcccttac 1200
gggctgggct acacacgtgc tacaatgggt gtgacagtgg gcagcgagac agcgatgtcg 1260
agctaacttc caaaagccat ctcagttcgg attgcactct gcaactcgag tgcatagaag 1320
tggaatcgct agtaatcgca gatcagcatg tgcggtgaat 1360

```

<210> 64

<211> 1288

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 64

```

tccaggaaac agctatgacc atgattacgc caagcttggg accgagctcg gatccactag 60
taacggccgc cagtgtgctg gaattcgccc ttcaggccta acacatgcaa gtcgagcgcc 120
ccgcaagggg agcggcagac ggggtgagtaa cgcgtgggaa tctacccatc cctacggaac 180
aactccggga aactggagct aataccgtat acgccttttg ggggaaagat ttatcgggga 240
tggtgagacc cgcgttggat tagctagtgg gtggggtaaa ggccctacaa ggcgacgac 300
catagctggg ctgagaggat gatcagccac attgggactg agacacggcc caaactccta 360
cgggaggcag cagtggggaa tattggacaa tgggcgcaag cctgatccag ccatgcccgc 420
gtgagtgatg aaggctcttag gattgtaaag ctctttcacc ggagaagata atgacggtat 480
ccggagaaga agccccggct aactttcgtg ccagcagccg cggtataacg aagggggcta 540
gcgttggtcg gaattactgg gcgtaaagcg cacgtaggcg gatatttaag tcaggggtga 600
aatcccagag ctcaactctg gaactgcctt tgatactggg tatcttgagt atggaagagg 660
taagtggaa 720
tgcgagtgtg gaggtgaaat tcgtagatat tcgcaggaa accagtggcg 720
aaggcggtt actggtccat tactgacgct gaggtgcgaa agcgtgggga gcaaacagga 780
ttagataccc tggtagtcca cgccgtaaac gatgaatgtt agccgtcggc aagtttactt 840
gtcgggtggc cagctaacgc attaaacatt ccgcctgggg agtacggtcg caagattaaa 900

```

```

actcaaagga attgacgggg gcccgacaaa gcggtggagc atgtgggtta attcgaagca 960
acgcgcagaa ccttaccagc ccttgacatg cccggacagc tacagagatg tagtggtccc 1020
ttcggggacc gggacacagg tgctgcatgg ctgtcgtcag ctctgtctgt gagatgttgg 1080
gttaagtccc gcaacgagcg caaccctcgc ccttagttgc cagcattcag ttgggcactc 1140
taaggggact gccggtgata agccgagagg aagtggggat gacgtcaagt cctnatggcc 1200
cttacgggct gggctacaca cgtgctacaa tgggtgggtga cagtgggcag cgaaggaacg 1260
atcccagact aatctccaaa agccatct 1288

```

<210> 65

<211> 1386

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 65

```

cgacggccag tgaattgtaa tacgactcac tataggcgga attgggccct ctagatgcat 60
gctcgagcgg ccgccagtgt gatggatata tgcagaattc gcccttcagg cctaacacat 120
gcaagtcgag cgggcgtagc aatacgtcag cggcagacgg gtgagtaacg cgtgggaaca 180
taccttttgg ttccgaacaa cacagggaaa cttgtgctaa taccggataa gcccttacgg 240
ggaaagattt atcgccgaaa gattggcccc cgtctgatta gctagttagt agggtaatgg 300
cctaccaagg cgacgatcag tagctggtct gagaggatga tcagccacat tgggactgag 360
acacggccca aactcctacg ggaggcagca gtggggaata ttggacaatg ggcgcaagcc 420
tgatccagcc atgccgcgtg agtgatgaag gccctagggt tgtaaagctc ttttgtgcgg 480
gaagataatg acggtaccgc aagaataagc cccggctaac ttctgtccag cagccgcggg 540
aatacgaagg gggctagcgt tgctcggaat cactgggcgt aaagggtgcg taggcgggtc 600
tttaagtcag gggtgaaatc ctggagctca actccagaac tgcccttgat actgaagatc 660
ttgagttcgg gagaggtagg tggaactgcg agtgtagagg tgaaattcgt agatattcgc 720
aagaacacca gtgggcgaag gcggctcact ggcccgatac tgacgctgag gcacgaaagc 780
gtggggagca aacaggatta gataccctgg tagtccacgc cgtaaacgat gaatgccagc 840
cgttagtggg tttactcact agtggcgtag ctaacgcttt aagcattccg cctggggagt 900
acggtcgcaa gattaaaact caaaggaatt gacggggggc cgcacaagcg gtggagcatg 960
tggtttaatt cgacgcaacg cgcagaacct taccagccct tgacatgtcc aggaccggtc 1020
gcagagatgt gaccttctct tcggagcctg gagcacaggg gctgcatggc tgtcgtcagc 1080
tcgtgtcgtg agatgttggg ttaagtcccg caacgagcgc aacccccgtc cttagttagt 1140
accatttagt tgagcactct aaggagactg ccggtgataa gccgcgagga aggtggggat 1200
gacgtcaagt cctcatggcc cttacgggct gggctacaca cgtgctacaa tggcggtgac 1260
aatgggacgc taaggggcaa cccttcgcaa atctcaaaaa gccgtctca gtteggattg 1320
ggctctgcaa ctcgagccca tgaagttaga atcgctagta atcgtggatc agcacgccac 1380
ggtgaa 1386

```

<210> 66

<211> 1223

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 66

```

agcggcagag ggtgagtaac gcgtgggaat ctacccatct ctacggaaca actccgggaa 60
actggagcta ataccgtata cgtccttcgg gagaaagatt tatcggagat ggatgagccc 120
gcgttggtatt agctagttgg tggggtaatg gcctaccaag gcgacgatcc atagctggtc 180
tgagaggatg atcagccaca ctgggactga gacacggccc agactcctac gggaggcagc 240
agtggggaat attggacaat gggcgaaagc ccgatccagc catgccgcgt gagtgatgaa 300
ggccctaggg ttgtaaagct ctttcaacgg tgaggataat gacggtaacc gtagaagaag 360
ccccggctaa cttcgtgcc a gcagccgagg taatacgaag ggggctagcg ttgttcggaa 420
ttactgggag taaagcgcac gtaggcggac tattaagtca ggggtgaaat cccggggctc 480
aaccgccgaa ctgcctttga tactggtagt ctcgagtcgg gaagagggtga gtggaattcc 540
gagtgtagag gtgaaattcg tagatattcg gaggaacacc agtggcgaag gcggctcact 600
ggtcgggtac tgacgctgag gtgcgaaagc gtggggagca aacaggatta gataccctgg 660
tagtcacgcg cgtaaacgat ggaagctagc cgttggcaag tttacttgct ggtggcgag 720
ctaacgcatt aagcttccc cctggggagt acggtcgcaa gattaaaact caaaggaatt 780
gacggggggc cgcacaagcg gtggagcatg tggtttaatt cgaagcaacg cgcagaacct 840
taccagccct tgacatccc gtcgcgggta ccagagatgg tatccttcag ttcggctgga 900
ccgggtgacg gtgctgcatg gctgtcgtca gctcgtgtcg tgagatgttg ggttaagtcc 960
cgcaacgagc gcaaccctcg cccttagttg ccagcattca gttgggcact ctaaggggac 1020
tgccgggtgat aagccgagag gaaggtgggg atgacgtcaa gtcctcatgg cccttacggg 1080
ctgggctaca cacgtgctac aatgggtggg acagtgggca gcgagaccgc gaggtcgagc 1140
taatctcaa aagccatctc agttcggatt gcactctgca actcgagtgc atgaagtgg 1200
aatcgctagt aatcgcggat cag                                     1223

```

<210> 67

<211> 1237

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 67

```

cccgaggagg agtggcagag ggtgagtaac gcgtgggaat ctaccctttt ctacggaaca 60
actgagggaa acttcagcta ataccgtata cggccgagag gcgaaagatt tatcggagaa 120
ggatgagccc gcgttggtatt agctagttgg tggggtaaaag gcctaccaag gcgacgatcc 180
atagctggtc tgagaggatg atcagccaca ctgggactga gacacggccc agactcctac 240
gggaggcagc agtggggaat attggacaat gggcgcaagc ctgatccagc catgccgcgt 300
gagtgatgaa ggccctaggg ttgtaaagct ctttcacggg tgaagataat gacggtaacc 360
ggagaagaag ccccggttaa cttcgtgcc a gcagccgagg taatacgaag ggggctagcg 420
ttgttcggat ttactgggag taaagcgcac gtaggcggac tattaagtca ggggtgaaat 480
cccggggctc aaccgccgaa ctgcctttga tactggtagt cttgagttcg aaagagggtga 540
gtggaattcc gagtgtagag gtgaaattcg tagatattcg gaggaacacc agtggcgaag 600
gcggctcact ggctcgatac tgacgctgag gtgcgaaagc gtggggagca aacaggatta 660
gataccctgg tagtcacgcg cgtaaacat gagagctagg cgtcgggcag tatactgttc 720
ggtggcgag ctaacgcatt aagctcttcg cctggggagt acggtcgcaa gattaaaact 780
caaaggaatt gacggggggc cgcacaagcg gtggagcatg tggtttaatt cgaagcaacg 840
cgcagaacct taccagccct tgacatccc atcgcgggta ccagagatgg tatccttcag 900

```

```

ttaggctgga tcggtgacag gtgctgcatg gctgtcgtca gctcgtgtcg tgagatgttg 960
ggttaagtcc cgcaacgagc gcaaccctcg cccttagttg ccatcattca gttgggcact 1020
ctaaggggac tgccggtgat aagccgagag gaaggtgggg atgacgtcaa gtcctcatgg 1080
cccttacggg ctgggctaca cacgtgctac aatgggtggc acagtgggca gcgagaccgc 1140
gaggtcgagc taatctccaa aagccatctc agttcggatt gcactctgca actcgagtgc 1200
atgaagttag aatcgctagt aatcgtggat cagaatg 1237

```

<210> 68

<211> 1346

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 68

```

acgacggggc agtgaattgt aatacgactc actatagggc gaattggggc ctctagatgc 60
atgctcgagc ggccgccagt gtgatggata tctgcagaat tcgcccttca ggcctaacac 120
atgcaagtcg aacggatccc ttcggattag tggcggacgg gtgagtaaca cgcggaacg 180
tgcccttttg ttcggaacaa cttagggaaa cttgagctaa taccggataa gccttttcgag 240
ggaaagattt atcgccattg gagcgggccc cgtaggatta gctagttggt gaggtaaaag 300
ctcaccaagg cgacgatcct tagctggtct gagaggatga tcagccacat tgggactgag 360
acacggccca aactcctacg ggaggcagca gtgggggaatc ttgcgcaatg ggcgaaaagc 420
tgacgcagcc atgccgcgtg aatgatgaag gtcttaggat tgtaaaattc tttcaccggg 480
gacgataatg acggtacccg gagaagaagc cccggctaac ttcgtgccag cagccgcggt 540
aatacgaagg gggctagcgt tgctcggaat tactgggcgt aaagggagcg taggcggata 600
gttttagtcag aggtgaaagc ccagggtcga accttggaat tgcccttgat actggctatc 660
ttgagtacgg aagaggtatg tggaaactcc agtgtagagg tgaaattcgt agatattcgg 720
aagaacacca gtggcgaaag cgacatactg gtccgttact gacgctgagg ctcgaaaagc 780
tggggagcaa acaggattag ataccctggg agtccacgct gtaaacgatg agtgctagtt 840
gtcggcatgc atgcatgtcg gtggcgagc taacgcatta agcactccgc ctggggagta 900
cggctcgcaag attaaaactc aaaggaattg acggggggccc gcacaagcgg tggagcatgt 960
ggtttaattc gaagcaacgc gcagaacctt accacctttt gacatgcccg gaccgctcca 1020
gagatggagc tttcccttcg gggactggga cacagggtgct gcatggctgt cgtcagctcg 1080
tgctgtgaga tggtgggtta agtcccgcga cgagcgcaac cctcgctatt agttgccatc 1140
aggtttggct gggcactcta ataggaccgc cgggtggtaag ccggaggaag gtggggatga 1200
cgtcaagtcc tcatggccct tacaaggtgg gctacacacg tgctacaatg gcgactacag 1260
agggctgcaa tcccgcgagg gggagccaat ccctaaaagt cgtctcagtt cggattgcac 1320
tctgcaactc gagtgcatag agttgg 1346

```

<210> 69

<211> 1500

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 69

```

acagctatga ccatgattac gccaaagcttg gtaccgagct cggatccact agtaacggcc 60
gccagtgtgc tgggaattcgc ctttcaggcc taacacatgc aagtcgaacg ccgtagcaat 120
acggagtggc agacgggtga gtaacacgtg ggaacgtgcc ctttggttcg gaacaacaca 180
gggaaacttg tgctaatacc gaataagccc ttacggggaa agatttatcg ccaaaggatc 240
ggcccgcgtc tgattagcta gttggtgggg taacggccca ccaaggctac gatcagtagc 300
tggtctgaga ggatgatcag ccacactggg actgagacac ggcccagact cctacgggag 360
gcagcagtta ggaatcttgg acaatgggcg caagcctgat ccagccatgc cgcgtgagtg 420
atgaaggcct tagggttgta aagctctttc agcggggaag ataatgacgg taccgcaga 480
agaagccccg gctaacttcg tgccagcagc cgcggtaata cgaagggggc tagcgttgct 540
cggaatcact gggcgtaaaag cgcacgtagg cggatcttta agtcaggggt gaaatcctgg 600
agctcaactc cagaactgcc ttgatactg gggatctcga gtccggaaga ggtgagtggg 660
actccgagtg tagaggtgaa attcgtagat attcgggaaga acaccagtgg cgaaggcggc 720
tcaactggcc ggtactgacg ctgaggtgcg aaagcgtggg gagcaaacag gattagatac 780
cctggtagtc cacgccgtaa acgatggatg ctagccgttg gcgggtttac tcgtcagtgg 840
cgcagctaac gcattaagca tcccgcttg ggagtacggg cgcaagatta aaactcaaag 900
gaattgacgg gggcccgcac aagcgggtgga gcatgtggtt caattcgaag caacgcgcag 960
aaccttacca gcccttgaca tgtcccgat ggacttcaga gatgaggtcc ttcagttcgg 1020
ctggcgggaa cacaggtgct gcatggctgt cgtcagctcg tgtcgtgaga tgttgggtta 1080
agtcccgcaa cgagcgcaac cctcgccctt agttgccatc atttagttgg gcactctaag 1140
gggactgccg gtgataagcc gcgaggaagg tggggatgac gtcaagtcct catggccctt 1200
acgggctggg ctacacacgt gctacaatgg cggtgacagt gggacgcaat ggagcaatcc 1260
tgcgcaaatc tcaaaaagcc gtctcagttc ggattggggg ctgcaactcg accccatgaa 1320
gtcggaatcg ctagtaatcg cagatcagca cgctgcggtg aatacgttcc cgggccttgt 1380
acacaccgcc caagggcgaa ttctgcagat atccatcaca ctggcgggcg ctcgagcatg 1440
catctagagg gcccaattcg ccctatagtg agtcgtatta caattcactg gccgtcgttt 1500

```

<210> 70

<211> 1113

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence : Organisme du sol

<400> 70

```

gagctaatac cgtataatga cttcgggtcca aagatttatc gcctgaggat gagcccgcg 60
cggattagct agttggtagg gtaaaagcct accaaggcga cgatccgtag ctggtctgag 120
aggatgatca gccacactgg gactgagaca cggcccagac tcctacggga ggcagcagt 180
gggaatattg gacaatgggc gcaagcctga tccagcaatg ccgcgtgagt gatgaaggcc 240
ttagggttgt aaagctcttt taccgggaa gataatgact gtaccgggag aataagcccc 300
ggctaactcc gtgccagcag ccgcggtaat acggaggggg ctagcgttgt tcggaattac 360
tgggcgtaaa gcgcacgtag gcggctttgt aagttagagg tgaaagcccc gggctcaact 420
ccggaattgc ctttaagact gcatcgctcg aattgtggag aggttaagtgg aattccgagt 480
gtagaggtga aattcgtaga tattcggaag aacaccagtg gcgaaggcga cttactggac 540
acatattgac gctgaggtgc gaaagcgtgg ggagcaaaca ggattagata ccctggtagt 600
ccacgccgta aacgatgatg actagctgtc ggggcgctta gcgtttcggg ggcgagccta 660
acgcgttaag tcatccgcct ggggagtagc gccgcaaggt taaactcaaa gaaattgacg 720
ggggcctgca caagcgggtg agcatgtggt ttaattcgaa gcaacgcgca gaaccttacc 780

```

```

agcgtttgac atgccaggac ggtttccaga gatggattcc ttcccttacg ggacctggac 840
acaggtgctg catggctgtc gtcagctcgt gtcgtgagat gttgggttaa gtcccgaac 900
gagcgcaacc ctctcttcta gttgctacca tttagttagg cactctagag aaactgccgg 960
tgataagccg gaggaagggt gggatgacgt caagtcctca tggcccttac gcgctgggct 1020
acacacgtgc tacaatggcg gtgacaacgg gcagcaaaact cgcgagagtg agcaaatccc 1080
gaaaagccgt ctcatgttcgg attgttctct gca 1113

```

<210> 71

<211> 1225

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 71

```

ggagcggcgg acgggtgagt aacgcgtggg aacgtgccct ttggtacgga acaactgagg 60
gaaacttcag ctaataccgt atgtgccctt cgggggaaag atttatcgcc attggagcgg 120
cccgcgttgg attaggtagt tgggtggggt aaggcctacc aagcctacga tccatagctg 180
gtctgagagg atgatcagcc acactgggac tgagacacgg ccagactcc tacgggaggc 240
agcagtaggg aatcttgccg aatgggcgaa agcctgacgc agccatgccg cgtgtatgat 300
gaaggtctta ggattgtaa atactttcac cggggaagat aatgacggta cccggagaag 360
aagccccggc taacttcgtg ccagcagccg cggtaatacg aagggggcta gcgttgctcg 420
gaattactgg gcgtaaaggg cgcgtaggcg gatatttaag tcgggggtga aagcccaggg 480
ctcaaccctg gaattgcctt cgatactgga tatcttgagt tcgggagagg tgagtggaat 540
gccgagtgtg gaggtgaaat tcgtagatat tcggcggaac accagtggcg aaggcgactc 600
actggccgga tactgacgct gaggcgcgaa agcgtgggga gcaaacagga ttagataccc 660
tggtagtcca cgctgtaaac gatgagtgt agttgtcggc atgcatgcat gtcggtgacg 720
cagctaacgc attaagcact ccgcctgggg agtacgggtc caagattaaa actcaaagga 780
attgacgggg gccgcacaaa gcggtggagc atgtggttta attcgaagca acgcgcagaa 840
ccttaccacc ttttgacatg ccctgatcgc tggagagatc cagttttccc ttcggggaca 900
gggacacagg tgctgcatgg ctgtcgtcag ctctgtcgt gagatgttgg gttaagtccc 960
gcaacgagcg caaccctcgc cattagttgc catcattaag ttgggcactc taatgggacc 1020
gccggtggta agccggagga aggtggggat gacgtcaagt cctcatggcc cttacggggg 1080
gggctacaca cgtgctacaa tggcgactac agaggggttg aaacctgcga aggggagcta 1140
atccctaaaa gtcgtctcag ttcggattgc actctgcaac tcgagtgcag gaagtcggaa 1200
tcgctagtaa tcgcgatca gcatg 1225

```

<210> 72

<211> 1286

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 72

```

atgattagta gcaataactaa tcgatgacga gcggcggacg ggtgagtaat acgtaggaac 60

```

```

ctgcccttaa gcggggggata actaagggaa acttttagcta ataccgcata aactcgagag 120
agaaaagctg cagcaatgtg gcacttgagg aggggcctgc gtcagattag ctagttagtg 180
aggtaatagc tcaccaaggc gatgatctgt aactggtctg agaggacgac cagtcacact 240
gggactgaga cacggcccag actcctacgg gaggcagcag tggggaatat tggacaatgg 300
gggcaaccct gatccagcga tgcgcgtgg gtgaagaagg ccttcggggt gttaaagcct 360
ttaggtcggg aagaaggtta gtagaggaaa tgctattaac ttgacggtac cgacagaata 420
agcaccggca aactctgtgc cagcagccgc ggtaatacag aggggtgcgag cgttaatcgg 480
atttactggg cgtaaagggc gcgtaggcgg tgagatgtgt gtgatgtgaa agccccaggc 540
tcaacctggg aagtgcacgc caaactgtct gactggagta tatgagaggg tggcggaatt 600
tccggtgtag cgggtgaaatg cgtagagatc ggaaggaaacg tcgatggcga aggcagccac 660
ctggcataat actgacgctg aggcgcgaaa gcgtggggat cgaacaggat tagataccct 720
ggtagtccac gctgtaaaact atgagtacta gatgttggtg ggggaacctg tcggtatcga 780
agctaacgcg ataagtattc cgctgggaa gtacggccgc aagggtgaaa ctcaaataaa 840
ttgacggggg ccgcacaaag cgggtggagca tgtggtttaa ttgatgcaa cgcgaagaac 900
cttacctacc cttgacatcc tgagaatctg gcttagtagc tggagtgcgc aaaggagctc 960
agagacaggt gctgcatggc tgtcgtcagc tcgtgttggt agatgttggg ttaagtccc 1020
taacgagcgc aacctttgcc cttagtgtgc atcatttagt tggggactct aaggggaccg 1080
ccagtgatga actggaggaa ggcggggacg acgtcaagtc atcatggcct ttatgggtag 1140
ggccacacac gtgtacaat ggggcgtacg gagggtcgca aaccgcgag ggggagctaa 1200
tctcataaag cgtctcgtag tccggattgg agtctgcaac tcgactccat gaagttggaa 1260
tcgctagtaa tcgcgaatca gcattg 1286

```

<210> 73

<211> 1288

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 73

```

cggggcaacc ctggcggcga gcggcgaacg ggtgagtaat gcatcggaac gtgtcctctt 60
gtgggggata accagtcgaa agactggcta ataccgcag agatcgaaag atgaaagcag 120
gggaccgcaa ggccttgccg gagaggagca gccgatgccg gattagctag ttggtgggg 180
aaaagcctac caaggcgacg atccgtagct ggtctgagag gacgaccagc cacactggga 240
ctgagacacg gccagactc ctacgggagg cagcagtggg gaattttgga cagtgggggc 300
aaccctgacg cagccatgcc gcgtgtgtga agaaggcctt cgggttgtaa agcactttcg 360
gacggaacga aatcgcgcg gttaatagtt cgcgtggatg acggtaccgt aagaagaagc 420
accggctaac tacgtgccag cagccgcggt aatacgtagg gtgcgagcgt taatcggaat 480
tactgggcgt aaagtgtgcg caggcggcct cgcaagtcga gtgtgaaatc cccgagctta 540
acttgggaat tgcgtcga actacggagc cggagtgtgg cagagggaagg tgggaattcca 600
cgtgtagcgg tgaaatgcgt agagatgtgg aggaacaccg atggcgaagg cggccttctg 660
ggccaacact gacgtcatg cagcaaagcg tggggagcaa acaggattag ataccctgg 720
agtccacgcc ctaaacgatg atgactagtt gttggaggag ttaaatacct tagtaacgca 780
gctaacgcgt gaagtcacgc gcctggggag tacggtcgca agattaaaac tcaaagggaat 840
tgacgggggc ccgcacaagc ggtggatgat gtggtttaat tcgatgcaac gcgaaaaacc 900
ttacctacce ttgacatgct aggaacgctg cagaaatgta gcggtgccc aaagggaacc 960
tagacacagg tgcgtcatgg ctgtcgtcag ctcgtgtcgt gagatgttgg gtttaagtccc 1020
gcaacgagcg caaccctgc cattagttgc tacattcagt tgagcactct aatgggactg 1080

```

```

ccggtgacaa accggaggaa ggtggggatg acgtcaagtc ctcattggccc ttatgggtag 1140
ggctacacac gtcatacaat ggcgcgtaca gagggttgcc aaccgcgag ggggagccaa 1200
tcccagaaag cgcgtcgtag tccggattgg agtctgcaac tcgactccca tgaagtcgga 1260
atcgctagta atcgcggtatc agcatgtc                                     1288

```

<210> 74

<211> 600

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 74

```

cgtgccagca gccgcggtaa tacgtagggtg gcaagcggtg tccggaatta ttgggcgtaa 60
agcgcgcgca ggtgggtttct taagtctgat gtgaaagccc acggcttaac cgtggagggt 120
cattggaac tgggagactt gagtgcagaa gaggaaagtga gaattccaag tgtagcgggtg 180
aaatgcgtag agatttggag gaacaccagt ggcgaaggcg actttctggt ctgcaactga 240
cgctgaggcg cgaaagcatg gggagcaaac aggattagat accctggtag tccatgccgt 300
aaacgatgag tgctaagtgt taggggggttt ccgcccctta gtgctgcagc taacgcatta 360
agcactccgc ctggggagta cgaccgcaag gttgaaactc aaaggaattg acgggggccc 420
gcacaagcgg tggagcatgt ggtttaattc gaagcaacgc gaagaacctt accaggtctt 480
gacatcccgat tganccgtct agagatagag ttttcccttc ggggacattg gtgacagggtg 540
gtgcatggtt gtcgtcagct cgtgtcgtga gatgttgggt taagtcccg aacgagcgca 600

```

<210> 75

<211> 601

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 75

```

cgtgccagca gccgcggtaa tacgtagggtg gcaagcggtg tccggaatta ttgggcgtaa 60
agcgcgcgca ggtgggtttct taagtctgat gtgaaagccc acggcttaac cgtggagggt 120
cattggaac tgggagactt gagtgcagaa gaggaaagtga gaattccaag tgtagcgggtg 180
aaatgcgtag agatttggag gaacaccagt ggcgaaggcg actttctggt ctgcaactga 240
cgctgaggcg cgaaagcatg gggagcaaac aggattagat accctggtag tccatgccgt 300
aaacgatgag tgctaagtgt taggggggttt ccgcccctta gtgctgagct aacgcattaa 360
gcactccgcc tggggagtag gaccgcaagg ttgaaactca aaggaattga cgggggcccc 420
cacaagcggg ggagcatgtg gtttaattcg aagcaacgcg aagaacctta ccaggtcttg 480
acatcccgat gacgtcttag agatagagtt ttcccttcgg ggacattggt gacaggtggt 540
gcatggttgt cgtcagctcg tgcgtgaga tggtgggtta agtcccgcaa cgagcgcacc 600
c                                                                 601

```

<210> 76

<211> 1236

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 76

```

tgccctgtag acggggataa cttcgggaaa ccggagctaa taccggataa tcctcttccc 60
cacatgggga agagttgaaa ggcgctttcg cgtcactaca ggatggggccc gcggtgcatt 120
agctagttag tagggtaacg gcctaccaag gcgacgatgc atagccgacc tgagaggggtg 180
atcgggcaca ttgggactga gacacggccc aaactcctac gggaggcagc agtaggggaat 240
cttcacaaat ggacgaaagt ctgatggagc aacgcccgcg gagtgatgaa ggttttcgga 300
tcgtaaaact ctgttgtaag ggaagaacca gtacgtcagg caatggacgt accttgacgg 360
taccttatta gaaagccacg gctaactacg tgccagcagc cgcggtaata cgtaggtggc 420
aagcgttgtc cggaattatt gggcgtaaag cgcgcgcagg tggtttctta agtctgatgt 480
gaaagcccac ggcttaaccg tggaggggtca ttggaaactg ggagacttga gtgcagaaga 540
ggaaagtgga attccaagtg tagcggcgaa atgcgtagag atttgaggga acaccagtgg 600
cgaaggcgac tttctggtct gcaactgacg ctgaggcgcg aaagcatggg gagcaaacag 660
gattagatac cctggtagtc catgctgtaa acgatgagtg ctaagtgtta gggggtttcc 720
gccccttagt gctgcagcta acgcattaag cactccgcct ggggagtagc accgcaagggt 780
tgaaactcaa aggaattgac gggggccccgc acaagcgggtg gagcatgtgg ttttaattcga 840
agcaacgcga agaaccctac caggtcttga catcccgatg atcgctctgg agatagaggtt 900
ttcccttcgg ggacattggt gacaggtggt gcatggttgt cgtcagctcg tgcgtgaga 960
tggtgggtta agtcccgcaa cgagcgcaac ccttaatctt agttgccatc atttagttgg 1020
gcactctaag gtgactgccg gtgataaacc ggaggaagggt ggggatgacg tcaaatcatc 1080
atgcccctta tgacctgggc tacacacgtg ctacaatgga cgggtacaaag agtcgctaac 1140
tcgcgagagt atgctaattc catagaaccg ttctcagttc ggattgtagg ctgcaactcg 1200
cctacatgaa gccggaatcg ctagtaatcg cggatc

```

1236

<210> 77

<211> 815

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 77

```

caagcgttgt ccggaattat tgggcgtaaa gagctcgtag gcggtttgtc gcgtctgctg 60
tgaaaactcg aggtctcaacc tcgggcttgc agtgggtacg ggcagactag agtgcggtag 120
gggtgactgg aattcctggt gtagcgggtg aatgcgcaga tatcaggagg aacaccgatg 180
gcgaaggcag gtcactgggc cgcaactgac gctgaggagc gaaagcatgg ggagcgaaca 240
ggattagata ccctggtagt ccattgccgt aacgttgggc actaggtgtg gggctcattc 300
cacgagttcc gtgccgcagc aaacgcatta agtgccccgc ctggggagta cggccgcaag 360
gcttaaaact caaagaaatt gacggggggc cgcacaagcg gcggagcatg cggattaatt 420
cgatgcaacg cgaagaacct taccaaggct tgacatacac cggaaacttc cagagatggg 480
tgccccgcaa ggtcgggtga caggtgggtg atggttgtcg tcagctcgtg tcgtgaagat 540
gttggtgtaa gtcccgaac gagegcgaacc ctcgtcctat gttgccagca cgtgatgggt 600

```

```

gggactcata ggagactgcc ggggtcaact cggaggaagg tggggatgac gtcaaatcat 660
catgcccctt atgtcttggg cttcacgcat gctacaatgg ccggtacaaa gggctgcat 720
accgcaaggt ggagcgaatc ccaaaaagcc ggtctcagtt cggattgggg tctgcaactc 780
gaccccatga agtcggagtc gctagtaatc gcaga 815

```

<210> 78

<211> 826

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 78

```

tcgtaggtgg cttgtcacgt cgggtgtgaa agcttggggc ttaactccag gtctgcattc 60
gatacgggct ggctagaggt aggtagggga gaacggaatt cctggtgtag cggtgaaatg 120
cgcagatatc aggaggaaca ccggtggcga aggcggttct ctgggcctta cctgacgctg 180
aggagcgaag gcgtggggag cgaacaggat tagataccct ggtagtccac gctgtaaacg 240
ttgggcgcta ggtgtgggga ccttccacgg tttccgcgcc gtagctaacg cattaagcgc 300
cccgcctggg gactacggcc gcaaggctaa aactcaaagg aattgacggg ggccccgaca 360
agcggcggag catgttgctt aattcgacgc aacgcgaaga accttacc aa ggcttgacat 420
cgccccgaaa gcttcagaga tggagccctc ttcggactgg gtgacagggt gtgcatggct 480
gtcgtcagct cgtgtcgtga gatgttgggt taagtccgc aacgagcgca acccttgctc 540
aatgttgcca gcaacatcct tcggggtggg tggggactca ttggagactg ccgggggtcaa 600
ctcggaggaa ggtggggacg acgtcaagtc atcatgcccc ttatgtcttg ggctgcaaac 660
atgctacaat ggccggtaca gaggggttgcg ataccgcaag gtggagcgaa tccctaaaag 720
ccggtctcag ttcggattgg ggtctgcaac tcgaccccat gaagtcggag tcgctagtaa 780
tcgcagatca gcaacgctgc ggtgaataacg tccccgggcc ttgtac 826

```

<210> 79

<211> 799

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 79

```

cgtaggcggg ttgtcgcgtc tgccgtgaaa gtccggggct caactccgga tctgcgggtg 60
gtacgggcag actagagtga tgtaggggag actggaattc ctggtgtagc ggtgaaatgc 120
gcagatatca ggaggaacac cgatggcgaa ggcagggtctc tgggcattaa ctgacgctga 180
ggagcgaag catggggagc gaacaggatt agataccctg gtagtccatg ccgtaaaccgt 240
tgggcactag gtgtggggga cattccacgt tttccgcgcc gtagctaacg cattaagtgc 300
cccgcctggg gactacggcc gcaaggctaa aactcaaagg aattgacggg ggccccgaca 360
agcggcggag catgaggatt aattcgatgc aacgcgaaga accttacc aa ggcttgacat 420
gaaccgaaa cacctggaaa caggtgcccc gcttgcggtc gggttacagg tgggtgcatg 480
ttgtcgtcag ctcgtgtcgt gagatgttgg gttaagtccc gcaacgagcg caaccctcgt 540
tctatgttgc cagcgcgtta tggcggggac tcataggaga ctgccggggg caactcggag 600

```

```

gaaggtgggg acgacgtcaa atcatcatgc cccttatgtc ttgggcttca cgcattgctac 660
aatggccggt acaaagggtt gcgatactgt gaggtggagc taatcccaaa aagccggtct 720
cagttcggat tggggctctgc aactcgaccc catgaagtcg gagtcgctag taatcgcaga 780
tcagcaacgc tgcggtgaa 799

```

<210> 80

<211> 1250

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 80

```

tgccagcttg ctggtggatt agtggcgaac gggtagtaaa cacgtgagta acctgccctt 60
aactctggga taagcctggg aaactgggtc taatgccgga tatgactcct catcgcatgg 120
tggtgggtgg aaagcttttt gtggttttgg atggactcgc ggcctatcag cttgttgggtg 180
aggtaatggc tcaccaaggc gacgacgggt agccggcctg agaggggtgac cggccacact 240
gggactgaga cacggcccag acttctacgg gaggcagcag tggggaatat tgcacaatgg 300
gcgaaagcct gatgcagcga cgcgcgctga gggatgacgg ccttcggggt gttaaactct 360
ttcagtaggg aagaagcgaa agtgacggta cctgcagaag aagcgccggc taactacgtg 420
ccagcagccg cggtaatatc tagggcgcaa gcgttatccg gaattattgg gcgtaaagag 480
ctcgtaggcg gtttgtcgcg tctgccgtga aagtccgggg ctcaactccg gatctgagg 540
gggtacgggc agactagagt gatgtagggg agactggaat tcctgggtga gcggtgaaat 600
gcgcagatat caggaggaac accgatggcg aaggcaggtc tctgggcatt aactgacgct 660
gaggagcgaa agcatgggga gcgaacagga ttagataccc tggtagtcca tgccgtaaac 720
gttgggcact aggtgtgggg gacattccac gttttccgcg ccgtagctaa cgcattaagt 780
gccccgcctg gggagtacgg ccgcaaggct aaaactcaaa ggaattgacg ggggcccgc 840
caagcgccgg agcatgcgga ttaattcgat gcaacgcgag gaaccttacc aaggcttgac 900
atgaaccgga aatacctgga aacagggtgcc ccgcttgcg tgggtttaca ggtggtgcat 960
ggttgccgtc agctcgtgtc gtgagatgtt ggggttaagtc ccgcaacgag cgcaaccctc 1020
gttctatgtt gccagcgcgt tatggcgggg actcatagga gactgccggg gtcaactcgg 1080
aggaaggtgg ggacgacgtc aaatcatcat gccccttatg tcttgggctt cagcattgct 1140
acaatggccg gtacaaaggg ttgcgatact gtgaggtgga gctaatacca aaaagccgg 1200
ctcagttcgg attggggtct gcaactcgac cccatgaagt cggagtcgct 1250

```

<210> 81

<211> 1210

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 81

```

cgctaatacc ggatacggcg cgagagtctt cggactttcg cgagaaagat tcgcaaggat 60
cactgaggga cgagcctgcg gcccatcagc tagttggtga ggtaagagct caccaaggct 120
aagacgggta gctgggtctga gaggatgatc agccacactg gaactgagac acggtccaga 180

```

```

ctcctacggg aggcagcagt ggggaatatt gcgcaatggg cgaaagcctg acgcagccac 240
gccgcgtgag cgatgagggc cttcgggtcg taaagctctg tggggagaga cgaataaggc 300
cgggtgaagag tcggccttga cggatatctc ttagcaagca ccggctaact ccgtgccagc 360
agccgcggtg atacggaggg tgcaaacggt gctcggaatc attgggctga aagcgcacgt 420
aggcggcgtg ataagttggg tgtgaaagcc ctgggctcaa cccaggaagt gcattcaaaa 480
ctgtcacgct tgaatctcgg aggggggtcag agaattcccg gtgtagaggt gaaattcgta 540
gatatcggga ggaataccag tggcgaaggc gctggcctgg acgaagattg acgctgaggt 600
gcgaaagcgc ggggagcaaa caggattaga taccctggta gtccgcgctg taaacgatga 660
gtgctagacg ggggaggtat tgacccttc gctgccgaag ctaacgcgtt aagcactccg 720
cctggggagt acggtcgcaa gactaaaact caaaggaatt gacggggggc cgacaagcg 780
gtggagcatg tggtttaatt cgacgcaacg cgcaaacct tacctgggtt aaatccgccg 840
gaacctggct gaaaggctgg ggtgccctcc ggggaatcgg tgagaagggt ctgcatggct 900
gtcgtcagct cgtgtcgtga gatgttgggt taagtcccg aacgagcgca acccctatcg 960
tcagttgcc aattaaggt gggaactctg gcgagactgc cgggtctaaac cggaggaagg 1020
tggggacgac gtcaagtcct catggccctt atgccagggt ctacacacgt gctacaatgg 1080
ctggtacaat gagccgcaaa accgcgaggt caagctaata tcaaaaaacc agtctcagtt 1140
cggatcggag tctgcaactc gactccgtga agctggaatc gctagtaatc gaagatcagc 1200
acgctttcgg                                     1210

```

<210> 82

<211> 1272

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 82

```

gatgccagct tgctgggtgga ttagtggcga acgggtgagt aacacgtgag taacctgccc 60
ttaactctgg gataagcctg ggaaactggg tctaataccg gatatgactc ctcatcgcat 120
gggtggggggg ggaaagcttt ttgtggtttt ggatggactc gcggcctatc agcttgttgg 180
tgaggtaatg gctcaccaag gcgacgacgg gtagccggcc tgagaggggtg accggccaca 240
ctgggactga gacacggccc agactcctac gggaggcagc agtggggaat attgcacaat 300
gggcgaaaagc ctgatgcagc gacgccgcgt gagggatgac ggcttcggg ttgtaaacct 360
ctttcagtag ggaagaagcg aaagtgcagg tacctgcaga agaagcgccg gctaactacg 420
tgccagcagc cgcggttaata cgtagggcgc aagcgttatc cggaattatt gggcgtaaaag 480
agctcgtagg cggtttgtcg cgtctgccgt gaaagtccgg ggctcaactc cggatctgcg 540
gtgggtacgg gcagactaga gtgatgtagg ggagactgga attcctgggtg tagcgggtgaa 600
atgcgcagat atcaggagga acaccgatgg cgaaggcagg tctctgggca ttaactgacg 660
ctgaggaacg aaagcatggg gagcgaacag gattagatac cctggtagtc catgccgtaa 720
acgttgggca ctagggtgtgg gggacattcc acgttttccg cgccgtagct aacgcattaa 780
gtgccccgcc tggggagtac ggccgcaagg ctaaaactca aaggaattga cggggggccc 840
cacaagcggc ggagcatgcg gattaattcg atgcaacgcg aagaacctta ccaaggcttg 900
acatgaaccg gaaataacctg gaaacagggtg ccccgcttgc ggtcgggtta cagggtggtgc 960
atggttgtcg tcagctcgtg tcgtgagatg ttgggttaag tcccgcaacg agcgcaacc 1020
tcgttctatg ttgccagcgc gttatggcgg ggactcatag gagactgccg ggggtcaactc 1080
ggaggaagggt ggggacgacg tcaaatcctc atgcccccta tgtcttgggc ttacgcgatg 1140
ctacaatggc cgggtacaaaag gggtgcgata ctgtgaggtg gagctgatcc caaaaagccg 1200
gtcccagttc ggattgggggt ctgcaactcg accccatgaa gtcggagtcg ctagtaatcg 1260

```

cagatcagca ac

1272

<210> 83

<211> 1247

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 83

```

tgttttagtag caatactaaa tgatgacgag cggcggacgg gtgaggaaca cgtaggaacc 60
tgcccaagag agggggacaa ccaagggaaa ctttggctaa taccgcataa tctctacgga 120
gaaaagttgc ccgtaagggg ggcgcttttg gaggggcctg cgtccgatta gttagttggg 180
gaggtaatag ctcaccaaga ctgtgatcgg taactggtct gagaggacga ccagtcacac 240
tgggactgag acacggccca gactcctacg ggaggcagca gtggggaatc ttggacaatg 300
ggggcaaccc tgatccagcg atgccgcgtg ggtgaagaag gccttcgggt tgtaaagccc 360
tttaggcggg gaagaaggat atgggatgaa taagcctgta ttttgacggg acccgagaa 420
taagcacggg caaactctgt gccagcagcc gcggaatac agaggggtgc agcgttaatc 480
ggatttactg ggcgtaaagg gcgcgtaggc gggtgtgtga gtgtgatgtg aaagccccgg 540
gctcaacctg ggaagtgcac cgcaaacgac acaactggag tatatgagag ggtggcggaa 600
tttccgggtg agcggtgaaa tgcgtagaga tcggaaggaa cgtcgatggc gaaggcagcc 660
acctggcata atactggcgc tgaggcgcg aagcgtgggg agcgaacagg attagatacc 720
ctggtagtca cgcccgtaaa cgatgagaac tagatgttgg agggggaacc cttcagtatc 780
gaagctaacg cgataagttc tccgcctggg aagtacagtc gcaagactga aactcaaaag 840
aattgacggg ggcccgcaca agcgggtggg catgtggttt aattcgatgc aacgcgaaga 900
accttacctg cccttgacat cctgcgaatc ttgccgagag gtgagagtgc cgcagggagc 960
gcagagacag gtgctgcatg gctgtcgtca gctcgtgttg tgagatgttg ggttaagtcc 1020
cgtaacgagc gcaacccttg tccttagttg ccatcattta gttggggact ctaaggagac 1080
cgccggtgat gaaccggagg aaggcgggga cgacgtcaag tcatcatggc ctttatgggt 1140
agggctacac acgtgctaca atggggcgta cagagggtcg ccaaccgcg agggggagcc 1200
aatctcttaa agcgtctcgt agtccggatt ggagtctgca actcgac 1247

```

<210> 84

<211> 1292

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 84

```

ggctcgcaag agcaaccggc gaacgggtgc gtaacacgtg aacaacctgc cctcgtgtgg 60
gggatagccg ggctaacgcc cgggtaatac cgcatacgtt ctctctgggg agtcctgggg 120
agaggaaagc tccggcgcac ggggaggggt tcgcgccta tcagctagtt ggcggggtaa 180
tggccacca aggcgacgac gggtagctgg tctgagagga tggccagcca cattgggact 240
gagagacggc ccagactcct acgggaggca gcagtgggga atcttgcgca atggccgaaa 300
ggctgacgca gcgacgccgc gtgtgggagg acgcctttcg ggggtgaaac cactgttgcc 360

```

```

cgggacgaac agcctctttc gagaggtctg acggtaccgg gtgaggaagc accggctaac 420
tccgtgccag cagccgcggt aatacggagg gtgcgagcgt tgtccggaat cattgggctg 480
aaagggcgcg taggtggccc ggtcagttcg tggtgaaagc gcggggctca accctgcgtc 540
ggccatgaat actgccgcgg ctggagcact gtagaggcag gcggaattcc ggggtgtagc 600
gtggaatgcg tagagatccg gaagaacacc ggtggcgaaag gcggcctgct gggcagtagc 660
tgacactgag gcgcgacagc gtggggagca aacaggatta gataccctgg tagtccacgc 720
cgtaaacgat gggcactagg cgcttggggg agcgaccccc cgagggccgg cgctaacgca 780
ttaagtgcgc cgcttgggga gtacggccgc aaggctgaaa ctcaaaggaa ttgacggggg 840
cccgcacaa ggggtggagca tgtggttta ttcgacgcaa cgcaagaac cttacctagg 900
cttgacatac acgggaaacc ggtcagaaac ggcggccct cttcggagcc cgtgcacagg 960
tgctgcatgg ctgtcgtcag ctctgtctgt gagatgttgg gttaagtccc gcaacgagcg 1020
caaccctgt ctctagtgtc cagcgcgtca tggcggggac tctagagaga ctgccggtgc 1080
caaaccggag gaaggtgggg atgacgtcaa gtcacatgg tccttacgtc tagggctaca 1140
cacgtgctac aatggcgggg acagagggtc gcgagccggc aacggcaagc caatccccga 1200
aaccgcgct cagttcggat tgtcgtctgc aactcgacgg catgaagctg gaatcgctag 1260
taatcgtgga tcagctacgc cacgtggaat ac                                     1292

```

<210> 85

<211> 1300

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 85

```

tcccttcggg agcaagtaca gcggcgaacg ggtgagtaac acgtaggtaa cctaccctgg 60
agactgggat aacctgccga aaggcgggct aataccagat aagaccacga gggctgcggc 120
ccttggggca aaaggtggcc tctacttgta agctaccact ccgggatggg cctgcgcgcc 180
attagctagt tggcggggta acggccacc aaggcagaga tggctagctg gtctgagagg 240
atggccagcc acacagggac tgagacacgg ccagactcc tacgggaggc agcagtgggg 300
aatattgcgc aatgggcgaa agcctgacgc agcgacgccc cgtgggtgat gaaggccttc 360
gggtcgtaaa gccctgtcaa gagggacgaa acctgtgcga cctaacaagt cggcaacctg 420
acggtacctc tgaaggaagc accggctaac tccgtgccag cagccgcggt aatacggagg 480
gtgcgagcgt tgttcggaat tactgggcgt aaagcgcgtg taggcggcct cttcagctctg 540
gtgtgaaagc ccggggctca accccggaag tgcattggat actgggaggc tggagtaccg 600
gagaggaggg tgggaattcct ggtgtagcgg tgaaatgcgt agatatcagg aggaacacct 660
gtggcgaaag cggccctctg gacggatact gacgctgaga cgcgaaagcg tggggagcaa 720
acaggattag ataccctggt agtccacgct gtaaacgatg ggcactaggt gttcggggta 780
ttgacccctc gagtgcgcga gctaacgcat taagtgcctc gcctggggaa tacggccgca 840
aggttaaaac tcaaaggaat tgacgggggc ccgcacaagc ggtggagcat gtggtttaat 900
tcgacgcaac gcgaagaacc ttacctgggc tagacaacat cggacagcct cagaaatgag 960
gtctccccgc aaggggcccgg tggttcagggt gctgcatggc tgtcgtcagc tcgtgtcgtg 1020
agatgttggg ttaagtcccg caacgagcgc aaccctgtc tctagttgct accattcagt 1080
tgagcactct agagagactg ccngtggtta aacgggagga aggtggggac gacgtcaagt 1140
cctcatggcc cttatgtcca gggctacaca cgtgctacaa tgggcgatac aaagggtgct 1200
gaaccgcgca ggggaagcca atcccaaaaa gtcgctctca gttcggattg gagtctgcaa 1260
ctcgactcca tgaaggcgga atcgctagta atcgcgatc                                     1300

```

<210> 86
 <211> 1186
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 86
 caatgggcag cggcggacgg gtgagtaaca cgtgggaatg tacctttcgg tgcggaacaa 60
 ctcagggaaa cttgagctaa tgccgcatac gcccttacgg ggaaagattt atcgccgaaa 120
 gatcagcccc cgttggtatta gctagttggg gaggtaatgg ccaccaagg cgacgatcca 180
 tagctggttt gagagaacga ccagcctcac tgggactgag acacggccca gactcctacg 240
 ggaggcagca gttgggaatc ttggacaatg ggggaaaccc tgatccagcc atgccgcgtg 300
 agtgatgaag gccttcgggt tgtaaaactc ttctgacggg gacgataatg acggtaccgg 360
 tagaagaagc tccggctaac ttctgtgccg cagccgcggg aatacgaagg gggctagcgt 420
 tgttcggaat tactgggctg aaagcgtgcg caggcggcta tccaagtcag tgggtgaaagc 480
 ccggagctca actccggaat tgccattgaa actgttttagc ttgagtacga gagaggtgag 540
 tggaataccc agtgtagagg tgaaattcgt agatattggg tagaacaccg gtggcgaagg 600
 cggctcactg gctcgttaact gacgctcagg cacgacagcg tggggatcaa acaggattag 660
 ataccctggg agtccacgcc gtaaacgatg aacgctagcc gttggatagc ttgctattca 720
 gtggcgcagc taacgcatta agcgttccgc ctggggagta cggccgcaag gttgagactc 780
 agaggaattg acggggggccc gcacaagcgg tggagcatgt ggtttaattc gacgcaacgc 840
 gcagaacctt accagggttt gacatcctgt gctcgccggg gaaagccggg tttcccgcga 900
 gggacgcaga gacaggtgct gcatggctgt cgtcagctcg tgctgtgaga tggtgggtta 960
 agtcccgcga cgagcgcaac cctcgccctt agttgccatc attcagttgg gcactctaga 1020
 gggaccgccc gcgacaagcc ggaggaaggg ggggatgacg tcaagtcccc atggccctta 1080
 caccctgggc tacacacgtg ctacaatggc ggtgacagtg ggcacgagct cgcgagagtc 1140
 agctaataccc aaaaaaccgt cccagttcag attgcactct gcaact 1186

<210> 87
 <211> 1454
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 87
 cgacggccag tgaattgtaa tacgactcac tatagggcga attgggccct ctagatgcat 60
 gctcgagcgg ccgccagtgt gatggatc tgcagaattc gcccttcagg cctaacacat 120
 gcaagtcgag cgagaaaggg cgcttcggcg cctgagtaca gcggcgcacg ggtgcgtaac 180
 acgtgggcaa tctgtccttg agatggggat aaccacagcg aagttgggct aataccgaat 240
 aagactacag gaggcaactc ccgtgggtta aggggtgctct ctgcggggag catgcgcttg 300
 aggaggagcc cgcggcctat cagctagttg gtagggtcac ggccctacaa ggcgaagacg 360
 ggtagctggg ctgagaggat gaccagccac acggggactg agacacggcc ccgactccta 420
 cgggaggcag cagtggggaa tattgggcaa tgggggaaac cctgaccag cgacgccgcg 480
 tgggtgatga aggcccttcgg gtcgtaaaagc cctgtcgggc ggaacgaagg ttctcacggc 540

```

aaatagccgt gagaggtgac ggtaccgccg aaggaagcac cggccaactc cgtgccagca 600
gccgcggtaa gacggaggggt gcaagcgttg ctcggaatca ctgggcgtaa aggggtgcgta 660
ggcggctctcg caagtctggc gtgaaagccc aaggctcagc cttggaagtg cgctcgaaac 720
tgcgaggctg gagtgccgga ggggagagtg gaattcccgg tgtagcgggtg aaatgcgtag 780
agatcgggag gaataccggg ggcgaaagcg actctctgga cggcaactga cgctgaggca 840
cgaaagcgtg gggagcaaac aggattagat accctggtag tccacgccgt aaacgatgga 900
cactaggtgt cgggggtatc cactccctcg gtgccgccgc taacgcagta agtgtcccgc 960
ctgggaagta cggtcgcaag attaaaactc aaaggaattg acggggggccc gcacaagcgg 1020
tggagcatgt ggttcaattc gatgcaacgc gaagaacctt acctgggttt gacatctggc 1080
gaatctctgg gaaaccagag agtgcccgcg ggggagcgcc aagacagggtg ctgcatggct 1140
gtcgtcagct cgtgccgtga ggtgttgggt taagtcccgc aacgagcgca acccttacct 1200
ttagttgccc ccgggtcaag ccgtggcact ccaagggaaac tgcccgtgtt aagcgggagg 1260
aaggtgggga cgacgtcaag tcatcatggc ctttatatcc agggctacac acgtgctaca 1320
atggctggga canagcgtgg ccaacgcgcg agcgggagct aatcgcaaaa cccagcctc 1380
agttcggatc ggagtctgca actcgactcc gtgaagctgg aatcgctagt aatcgcggtat 1440
cagcatgccg cgggt                                     1454

```

<210> 88

<211> 1307

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 88

```

cccttcgggg agcgagtaca gcggcgaacg ggtgagtaac acgtaggtaa cctaccctgg 60
tgactgggat aacttgccga aaggcgggct aataccagat aagaccacga gggctgcggc 120
ctttggggta aaagatggcc tctgcttgca tgetatcacg ccgggatggg cctgcgcgcc 180
attagctagt tggtaggta acggctcacc aaggcagaga tggctagctg gtctgagagg 240
atggccagcc acactgggac tgagacacgg cccagactcc tacgggaggg agcagtgggg 300
aatattgcgc aatgggcgaa agcctgacgc agcgacgccg cgtgggtgat gaaggccttc 360
gggtcgtaaa gccctgtcaa gagggacgaa acctcgccga cccaatacgt cggcgacctg 420
acggtacctc tgaaggaagc accggctaac tccgtgccag cagccgcggg aatacggagg 480
gtgcaagcgt tgttcggaat cactgggcgt aaagcgctg taggcggcct tcttagtctg 540
gtgtgaaagc ccggggctca accccggaag agcattggat actggaaggc tggagtaccg 600
gagaggaggg tggaaattcct ggtgtagcgg tgaaatgcgt agatatcagg aggaacaccg 660
gtggcgaagg cggccctctg gacggatact gacgctgaga cgcgacagcg tggggagcaa 720
acaggattag ataccctggg agtccacgcc gtaaaccgat ggtactaggt gttcggggta 780
ttgacccctc gagtgccgca gctaacgcat taagtacccc gcctggggac tacggccgca 840
aggctaaaac tcaaaggaat tgacgggggc ccgcacaagc ggtggagcat gtggtttaat 900
tcgacgcaac gcgaagaacc ttacctgggc tagacaacac tggacagccc cagaaatggg 960
gtcttcccgc aagggaactgg tggttcaggt gctgcatggc tgtcgtcagc tctgtctgtg 1020
agatgttggg ttaagtcccg caacgagcgc aaccctgtc tctagttgct accattaagt 1080
tgagcactct agagagactg cccgtgttaa acgggaggaa ggtggggagc acgtcaagtc 1140
ctcatggccc ttatgtccag ggctacacac gtgctacaat ggacagtaca aagggctgcg 1200
aaccctgtgag ggggagccaa tcccaaaaag ctgttctcag ttcggattgg agtctgcaac 1260
tcgactccat gaaggcgga tccgtagtaa tccgggatca gcatgcc                                     1307

```


<210> 89
 <211> 1305
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 89
 gggagcaatc cccaagtaga gcggcgaacg ggtgagtaac gcgtgggtaa tctgcctccg 60
 agtggggaac aacatcgga aactggtgct aataccgcat aacatcggtg ggtcttcgga 120
 tctgacgac aaagccgggg accgcaaggc ctggcgcttg gagaggagcc cgcgtccgat 180
 tagctagttg gtggggtaat ggcccaccaa ggcttcgac ggtagccggc ctgagagggc 240
 ggacggccac actgggactg agacacggcc cagactccta cgggaggcag cagtggggaa 300
 tttttcgcaa tgggcgaaag cctgacgaag caacgccgcg tggaggatga gggccttcg 360
 gtcgtaaact cctgtcgacc gggacgaaag taggatggcc taatacgccg atctattgac 420
 tgtaccggtg gaggaagcca cggctaactc tgtgccagca gccgcggtaa tacagaggtg 480
 gcaagcggtg ttcggaatta ctgggcgtaa agggcgcgta ggcggcttg ttcagtcccg 540
 gtgaaatccc tcggctcaac tgaggaactg cacgggaaac tgcctggctt gagttcgga 600
 gagggaagtg gaattccggg tgtagcgggtg aaatgcgtag atatccggag gaacaccggt 660
 ggcgaaggcg gcttcctgga ccgacactga cgctgaggcg cgaaagctag gggagcaaac 720
 gggattagat accccggtag tcctagctgt aaacgatgag tgctgggtgt agggggatc 780
 aacccccct gtgccgaagc taacgcatta agcactccgc ctggggagta cggctcgcaag 840
 gctgaaactc aaaggaattg acggggggcc gcacaagcgg tggagcatgt gggtcaattc 900
 gacgcaacgc gaagaacctt accgggggtt gaactgtacg ggacagctct agagatagag 960
 tcttccttcg ggaccgtac agaggtgctg catggctgtc gtcagctcgt gtcgtgagat 1020
 gttgggttaa gtcccgaac gagegcaacc cttgcctcct gttgccatca ggtaaagctg 1080
 ggcactctgg agagactgcc ggtgataaac cggaggaagg tggggatgac gtcaagtcct 1140
 catggccttt atgccccggg ctacacacgt gctacaatgg ccggtacaaa gggctcgaaa 1200
 accgcgaggt ggagctaata ccaaaaagcc ggtcccagtt cggattgcag tctgcaactc 1260
 gactgcatga agttggaatc gctagtaatc gcggatcagc atgcc 1305

<210> 90
 <211> 1299
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 90
 gggctttcgg gtctgagta aagtggcgaa cgggtgagta acgcgtaggt aacctgacct 60
 cgagtgtgga ataacctggc gaaagccggg ctaataccgc atgacgtctt cgggtcttcg 120
 gacttgagga ccaaagggtg cgagctttga gcgctgtcgc tcgagaaggg gctgcgtcc 180
 cattagctag ttgggtgggt gatggcctac caaggcgacg atgggtagcc gggctgagag 240
 gctgtccggc cacactggaa ccgagacacg gtccagactc ctacgggagg cagcagtggg 300
 gaatcttgcg caatggggga aacctgacg caacgacgcc gcgtgggcga tgaaggcctt 360
 cgggtcgtaa agccctgtcg agcgggacga accgtgcgag ctctaacata gctcgtgcct 420

```

gacggtaccg ctagaggaag ccccggttaa ctccgtgccg gcagccgcgg taatacggag 480
ggggctagcg ttattcggaa ttattgggcg taaagggcgt gtaggcggct ctgtgtgtcc 540
catgtgaaag ccctcggctc aaccggggaa ctgcatggga aactgcggag cttgagtccg 600
ggagaggtga gtggaattcc cagtgtagcg gtgaaatgcg tagatattgg gaggaacacc 660
agtggcgaag gcggctcact ggaccggtac tgacgtgag acgcgaaagc caggggagca 720
aacgggatta gataccccgg tagtcctggc tgtaaacgat gagcacttgg tgtggcgggt 780
atcgaccctt gccgtgctga agctaacgca ttaagtgtc cgctgggga gtacggccgc 840
aaggctgaaa ctcaaaggaa ttgacggggg ccgcacaag cggtgagca tgtggttcaa 900
ttcgacgcaa cgcaagaac cttacctggg tttgaactgc aggtgacagc ccctgaaagg 960
gggtcttctt tcgggacacc tgtagaggtg ccgcatggct gtcgtcagct cgtgtcgtga 1020
gatgttgggt taagtccgc aacgagcgca acccctact ctagttgccg gcggctcggc 1080
cgggaactct agggggaccg ccggtgataa accggaggaa ggtggggatg acgtcaagtc 1140
ctcatggcct ttatgtccag ggctacacac gtgctacaac ggacggtaca aagggtgcg 1200
aaggcgcgag ccggagccaa tcccaaaaag ccgttctcca gtgcggattg cagtctgcaa 1260
ctcgactgca tgaagggtga atcgctagta atcgcggt 1299

```

<210> 91

<211> 1296

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 91

```

atgtctggta gcaataccag atgatggcaa gtggcggacg ggtgagtaat acgtagggat 60
ctgcccagaa gagggggaca acccggggaa actcgggcta ataccgcata ctattctgag 120
gaagaaagct tggcgcaagc caggcgcttt tggaggaacc tacgtccgat tagctagtgt 180
gtgaggtaaa ggctcaccaa ggagagatc ggtagctggt ctgagaggat gaccagccac 240
actgggactg agacacggcc cagactccta cgggaggcag cagtggggaa tattggacaa 300
tgggggcaac cctgatccag cgatgccgcg tgtgtgaaga aggccttcgg gttgtaaagc 360
acttttagtt gggaagaagt aatgtttttt aatagagagc attgttgacg gtacccaaag 420
aataagcacc ggctaactct gtgccagcag ccgcggtaat acagagggtg caagcggttaa 480
tcggagttac tgggcgtaaa gggcgcgtag gcggtgttgc aagtgagatg tgaaatccct 540
gggcttaacc taggaaccgc attttagact gcaatgctag agtacagtag agggtagtgg 600
aatctccggg ttagcgggtg aatgcgtaga gatcggaagg aacaccagtg gcgaaggcga 660
ctacctggac tgacactgac gctgaggcgc gagagcgtgg ggagcaaaca ggattagata 720
ccctggtagt ccacgctgta aacgatgaga actagatgtt ggtgcgcgcg agcgcacaaag 780
tatcgaagct aacgcgataa gttctccgcc tggggagtag ggccgcaagg ttaaaactca 840
aaggaattga cgggggcccc cacaagcggg ggagcatgtg gtttaattcg atgcaacgcg 900
aggaacctta cctacccttg acatccacag aatttgatag agatatcgaa gtgccgaaag 960
gaactgtgag acagggtgct catggctgtc gtcagctcgt gttgtgagat gttgggttaa 1020
gtcccgtaac gagcgcaacc cttatcctta gttgccaaaca cgtaatggtg gggactctaa 1080
ggagactgcc ggtgaagaac cggaggaagg tggggacgac gtcaagtcac catggccttt 1140
atgggtaggg ctacacacgt gctacaatgg ggcgtacaga gggttgccaa cctgcgaagg 1200
ggagccaatc ccggaaagcg cctcgtagtc cagattgaag tctgcaactc gacttcatga 1260
agtcggaatc gctagtaatc gcgaatcaga acgtcc 1296

```

<210> 92
 <211> 1250
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 92
 gtctggtagc aataccagat gatggcaagt ggcggacggg tgagtaatac gtagggatct 60
 gcccagaaga gggggacaac ccggggaaac tcgggctaata accgcatact attctgagga 120
 aaaaagcttg gcgcaagcca ggcgcttttg gaggaacctt cgtccgatta gctagttggt 180
 gaggtaaagg ctcaccaagg cagagatcgg tagctggtct gagaggatga ccagccacac 240
 tgggactgag acacggccca gactcctacg ggaggcagca gtggggaata ttggacaatg 300
 ggggcaaccc tgatccagcg atgccgcgtg tgtgaagaag gccttcgggt tgtaaagcac 360
 tttagttggg gaagaagtaa tgttttttaa tagagagcat tgttgacggg acccaaagaa 420
 taagcaccgg ctaactctgt gccagcagcc gcggtataac agagggtgca agcggttaac 480
 ggagttactg ggcgtaaagg gcgcgtaggc ggtgttgcaa gtgagatgtg aaatccctgg 540
 gcttaacctt ggaaccgcat tttagactgc aatgctagag tacagtagag ggtagtggaa 600
 tttccggtgt agcgggtgaaa tgcgtagaga tcggaaggaa caccagtggc gaaggcgact 660
 acctggactg aactgacgc tgaggcgca gagcgtgggg agcaaacagg attagatacc 720
 ctggtagtc acgctgtaaa cgatgagaac tagatgttgg tgcgcgcgag cgcacaagta 780
 tcgaagctaa cgcgataagt tctccgctg gggagtacgg ccgcaagggt aaaactcaaa 840
 ggaattgacg ggggcccgc caagcgggtg agcatgtggt ttaattcgat gcaacgcgaa 900
 gaaccttacc tacccttgac atccacagaa tttgatagag atatcgaagt gccgaaagga 960
 actgtgagac aggtgctgca tggctgtcgt cagctcgtgt tgtgagatgt tgggttaagt 1020
 cccgtaacgg gcgcaaccct tatccttagt tgccaacacg taatggtggg gactctaagg 1080
 agactgccgg tgaagaaccg gaggaagggt gggacgacgt caagtcatca tggcctttat 1140
 gggtagggct acacacgtgc tacaatgggg cgtacagagg gttgccaacc tgccaagggg 1200
 agccaatccc ggaaagcgcc tcgtagtcca gattgaagtc tgcaactcga 1250

<210> 93
 <211> 1545
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 93
 ccaggaaaca gctatgacca tgattacgcc aagcttggtt ccgagctcgg atccactagt 60
 aacggccgcc agtgtgctgg aattcgccct tcaggcctaa cacatgcaag tcgaacggca 120
 gcacagggga gcttgctccc tgggtggcga gtggcgagc ggtgaggaat acatcggaat 180
 ctgcccagtc gtgggggata acctcgggaa accgggacta ataccgcata cgaccttagg 240
 gtgaaagcgg aggaccgcaa ggcttcgcgc gattggatga gccgatgtcg gattagcttg 300
 ttggcggggg aacggcccac caaggcgacg atccgtagct ggtctgagag gatgatcagc 360
 cacactggaa ctgagacacg gtccagactc ctacgggagg cagcagtggg gaatattgga 420
 caatgggcgc aagcctgac cagccatgcc gcgtgagtga agaaggcctt cgggttgtaa 480
 agctcttttg tccggaaaga aaagctttcg gttaataccc ggaagtcctg acggtaccgg 540

```

aagaataagc accggctaac ttcgtgccag cagccgcggt aatacgaagg gtgcaagcgt 600
tactcggaat tactgggcgt aaagcgtgcg taggtgggtt gttaagtctg atgtgaaagc 660
cctgggctca acctgggaat tgcactggat actggcaggc tagagtgcgg tagaggatgg 720
cggaattccc ggtgtagcag tgaaatgcgt agagatcggg aggaacatct gtggcgaaag 780
cggccatctg gaccagcact gacactgagg cacgaaagcg tggggagcaa acaggattag 840
ataccctggg agtccacgcc ctaaacgatg cgaactggat gttgggagca actaggctct 900
cagtatcgaa gctaacgcgt taagttcgcc gcctggggag tacggtcgca agactgaaac 960
tcaaaggaat tgacgggggc ccgcacaagc ggtggagtat gtggtttaat tcgatgcaac 1020
gcgaagaacc ttacctggcc ttgacatcca cggaaacttac cagagatggg ttggtgcctt 1080
cggnaaccgt gagacagggt ctgcatggct gtcgtcagct cgtgtcgtga gatgttgggt 1140
taagtcccg cccggtgaca aaccggagga aggtggggat gacgtcaagt catcatggcc 1260
cttacggcca gggctacaca cgtactacaa tggtcggtac agagggttgc aaagccgcga 1320
ggtagagcca atcccagaaa accgatccca gtccggatcg aagtctgcaa ctcgacttcg 1380
tgaagtcgga atcgctagta atcgcggtac agaagccgcg ggtgaatacg ttcccggggc 1440
ttgtacacac cgccccaggg cgaattctgc agatatccat cacactggcg gccgctcgag 1500
catgcatcta gagggcccaa ttcgccctat agtgagtcgt attac 1545

```

<210> 94

<211> 1549

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 94

```

ttttaaacgg acggccagtg aattgtaata cgactcacta tagggcgaaat tgggccctct 60
agatgcatgc tcgagcggcc gccagtgtga tggatatctg cagaattcgc ccttcaggcc 120
taacacatgc aagtcgagcg gcagcgcggg gcaacctggc ggcgagcggc ggacgggtga 180
ggaatgcatc ggaatctacc ctgtcgtggg ggataacgta gggaaactta cgctaatacc 240
gcatacgacc gagaggtgaa agtgggggac cgcaaggcct cacgcgtag gatgagccga 300
tgccggatta gctagttggg gaggtaaagg ctcaccaagg cgacgatccg tagctggtct 360
gagaggatga tcagccacat tgggactgag acacggccca aactcctacg ggaggcagca 420
gtggggaata ttggacaatg ggcgcaagcc tgatccagcc atgccgcgtg tgtgaagaag 480
gccttcgggt tgtaaagcac ttttgttcgg gaagaaatcg tgcgggttaa taccagtag 540
ggatgacggg accgaaagaa taagcaccgg ctaacttcgt gccagcagcc gcggtataac 600
gaaggggtgca agcgttactc ggaatcactg ggcgtaaagc gtgcgtaggc ggttgggttaa 660
gtctgctgtg aaagccctgg gctcaacctg ggaactgcag tggatactgg ccagctagag 720
tgtgatagag gatggtggaa ttcccgggtg agcggtgaaa tgcgtagaga tcgggaggaa 780
caccagtggc gaaggcggcc atctggatca acactgacgc tgaggcacga aagcgtgggg 840
agcaaacagg attagatacc ctggtagtcc acgccctaaa cgatgcgaac tggacgttgg 900
gagcaacttg gctctcagtg tcgaagctaa cgcgctaagt tcgccgcctg gggagtacgg 960
tcgcaagact gaaactcaaa ggaattgacg ggggcccgcg caagcgggtg agtatgtgg 1020
ttaattcgat gcaacgcgaa gaaccttacc tggccttgac atccacggaa cttaccagag 1080
atggtttggg gccttcggaa ccgtgagaca ggtgctgcat ggctgtcgtc agctcgtgtc 1140
gtgagatgtt ggggttaagtc ccgcaacgag cgcaaccctt gtccttagtt gccagcacgt 1200
aatggtggga actctaagga gactgccggg gacaaaccgg aggaagggtg ggatgacgtc 1260
aagtcacat gcccccttac gccagggcta cacacgtact acaatgggtc gtacaagagg 1320

```

```

gttgcaaagc ccgcgaggta gagccaatcc cagaaaaccc gatcccagtc ccggatcgaa 1380
gtctgcaact cgacttcgtg aagtcggaat cgctagtaat cgcggatcag aatgccgcgg 1440
tgaatacgtt cccgggcctt gtacacaccg cccaagggcg aattccagca cactggcggc 1500
cgttactagt ggatccgagc tcggtaccaa gcttggcgta atcatggtc 1549

```

<210> 95

<211> 1276

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 95

```

ctggcggcga gcgcgggacg ggtgaggaat acatcggaat ctaccagtc gtgggggata 60
acgtagggaa acttacgcta ataccgcata cgacctgagg gtgaaagcag gggatcgcaa 120
gaccttgccg gattggatga gccgatgtcc gattagctag ttggtgaggt aaaggctcac 180
caaggcgacg atcggtagct ggtctgagag ggtgatcagc cacactggaa ctgagacacg 240
gtccagactc ctacgggagg cagcagtggg gaattattga caatgggcgc aagcctgac 300
cagccatgcc gcgtgtgtga agaaggcctt cgggttgtaa agcacttttg ttcgggaaga 360
aatcttccga gttaatacct cgggaggatg acggtaccgg aagaataagc accggctaac 420
ttcgtgccag cagccgcggg aatacgaagg gtgcaagcgt tactcggaat tactgggcgt 480
aaagcgtgcg taggtggttc gttaagtctg ccgtgaaagc cccgggctca acctgggaat 540
tgcggtggat actggcgac tagagtgcgg tagagggtgg tggaattccc ggtgtagcag 600
tgaaatgctg agagatcggg aggaacatct gtggcgaagc ggccacctgg accagcactg 660
acactgaggc acgaaagcgt ggggagcaaa caggattaga taccctggta gtccacgccc 720
taaacgatgc gaactggacg ttgggagcaa ctaggctctc agtgtcgaag ctaacgcgtt 780
aagttcgccg cctggggagt acggtcgcaa gactgaaact caaaggaatt gacgggggcc 840
cgcacaagcg gtggagtgtg tggtttaatt cgatgcaacg cgaagaacct tacctggcct 900
tgacatccac ggaatccttt agagatagag gagtgccttc gggaaccgtg agacaggtgc 960
tgcattggctg tcgtcagctc gtgtcgtgag atgttgggtt aagtcgccga acgagcgcaa 1020
cccttgcct tagttgccag cgcgtaatgg cgggaactct aaggagactg ccggtgacaa 1080
accggaggaa ggtggggatg acgtcaagtc atcatggccc ttacggccag ggctacacac 1140
gtactacaat ggtggggaca gagggtcgcg aagccgcgag gtggagccaa tccagaaac 1200
cccatcctag tccggatcgg agtctgcaac tcgactccgt gaagtcgga tgcctagtaa 1260
tcgcggtcag catgcc 1276

```

<210> 96

<211> 1306

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 96

```

cagggatcag tagagtggca aacgggtgag taacgcgtgg gcgacctacc ttcgagtggg 60
ggataacctt ccgaaaggag ggctaatacc gcatgacatc ccgtgtttgg atacacggac 120

```

```

atcaaagccg gggatcgcaa gacctggcgc ttggagaggg gcccgcgtec gattagctag 180
ttggtgaggt cacggctcac caaggctccg atcggtatcc ggcctgagag ggcggacgga 240
cacactggga ctgagacacg gccagactc ctacgggagg cagcagtggg gaattgttcg 300
caatgggcgc aagcctgacg acgcaacgcc gcgtggagga tgaagacctt cgggtcgtaa 360
actcctttcg accgagatga agacccgcgc gcctaatacg ccggcggatt gacagtatcg 420
agggaaagaag ccccggttaa ctccgtgccg gcagccgcgc taatacgggg ggggcaagcg 480
ttgttcggaa ttactgggcg taaaggggtc gtaggtggct cgctaagtca gacgtgaaat 540
ccctcagctc aactggggaa ctgcgtctga gactggcaag cttgagtgcg ggagaggaa 600
gcggaattcc aggtgtagcg gtgaaatgcg tagatatctg gaggaacacc ggtggcgaa 660
gcggcgttct ggactgeaac tgacactgag gaacgaaagc taggggagca aacgggatta 720
gatacccccg tagtcctagc cctaaacgat gaatgcttgg tgtggcgggt atcgatccct 780
gccgtgccgc agttaacgcg ataagcatte cgcctgggga gtacggtcgc aaggctgaaa 840
ctcaaaggaa ttgacggggg cccgcacaag cgggtggagca tgtggttcaa ttcgacgcaa 900
cgcgaaagaac cttacctagg ctggaagtgc agatgaccat cggtgaaagc cgactttcgc 960
aagaacatct gtagaggtgc tgcattggctg tcgtcagctc gtgtcgtgag atgttgggtt 1020
aagtcccgca acgagcgcaa cccttgtttc ctgttgccat caggttaagc tgggcactct 1080
ggagagactg ccggtgacaa accggaggaa ggtggggatg acgtcaagtc agcatggcct 1140
ttatgtctag ggctacacac gtgctacaat ggccggtaca aagcgtgca aaccgcgag 1200
ggtgagccaa tcgcagaaag ccggtctcag ttcggatagc aggtgcaac tcgcctgctt 1260
gaagtggaa tcgctagtaa tcgcggatca gcatgccgcg gtgaat 1306

```

<210> 97

<211> 1300

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 97

```

cccgagggt gagtagatgg caaacgggtg agtaacacgt gggtgacctg cctcagagtg 60
ggggataacg acccgaaagg gtcgctaata ccgcataaca tcctgtcttt ggatagacgg 120
agatcaaagc cggggatcgc aagacctggc gcttagagag gggcccgcgc ccgattagct 180
agttggtgag gtaacggctc accaaggcaa cgatcggtat ccggcctgag agggcggacg 240
gacacactgg gactgagaca cggcccagac tcctacggga ggcagcagtg gggaaattgtt 300
cgcaatgggc gcaagcctga cgacgcaacg ccgcgtggag gatgaagatc ttcgggtcgt 360
aaactccttt cgatcgggaa gaacgcctct ggtgtgaaca ccatcagagg gtgacggtac 420
cgagagaaga agccccggct aactctgtgc cagcagccgc ggtaatacag ggggggcaag 480
cgttggttcg aattactggg cgtaaagggc tcgtaggcgc ccggctaagt ccgacgtgaa 540
atccccaggc ttaacctggg aactgcgtcg gatactggcg ggcttgaatc cgggagaggg 600
atgcggaatt ccagggtgtag cggtgaaatg cgtagatata tggaggaaaca ccggtggcga 660
aggcggcatc ctggaccggt attgacgctg aatagcgaaa gccaggggag caaacgggat 720
tagatacccc ggtagtcttg gccctaaacg atgaatgttt ggtgtggcgc gtatcgatcc 780
ctgcgtgcc gaagctaacg cattaacat tccgcctggg gactacggtc gcaaggctga 840
aactcaaagg aattgacggg ggcccgcaca agcgggtggag catgtggttc aattcgacgc 900
aacgcgaaga acctatccca ggctcgaacg gcattggaca tccggcgaaa gccggctccc 960
gcaagggccc atgtcgagggt gctgcattggc tgcgtcagc tcgtgtcgtg agatgttggg 1020
ttaagtcccg caacgagcgc aaccctgtc cgctgttgcc atcacgttat ggtgggcact 1080
ctgcggagac tgccgggtgat aaaccggagg aagggtgggga tgacgtcaag tcagcatggc 1140

```

```

ctttatgtct ggggctacac acgtgctaca atggccggta caaaccggtg cgatctcgca 1200
agagtgagct aatcggagaa agccgggtctc agttcggatt gcaggctgca actcgccctgc 1260
atgaagttgg aatcgctagt aatcgcggtat cagcacgccg 1300

```

<210> 98

<211> 1233

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 98

```

acggagcggc agacgggaga gtaacacgtg ggaacgtgcc ctttggttcg gaacaacaca 60
gggaaacttg tgctaatacc ggataagccc ttacggggaa agatttatcg ccaaaggatc 120
ggcccgcgtc tgattagcta gttggtgagg taacggctca ccaaggcgac gatcagtagc 180
tggtctgaga ggatgatcag cctcactggg actgagacac ggcccagact cctacgggag 240
gcagcagtgg ggaatattgg acaatgggcg caagcctgat ccagccatgc cgcgtggatg 300
atgaaggccc tagggttgta aagtcctttc ggcggggaag ataatgacgg taccgcgaga 360
agaagccccg gctaacttcg tgccagcagc cgcggtaata cgaagggggc tagcgttgct 420
cggaatcact gggcngtaaa gcgcacgtag gcggtctttt aagtcagggg tgaaatcctg 480
gagctcaact ccagaactgc ctttgatact gagaagcttg agtccgggag aggtgagtgg 540
aactgcgagt gtagaggtga aattcgtaga tattcgcaag aacaccagtg gcgaaggcgg 600
ctcactggcc cggtagtgac gctgaggtgc gaaagcgtgg ggagcaaaca ggattagata 660
ccctggtagt ccacgctgta aacgatggat gctagccgtt gtcgggttta ctcgtcagtg 720
gcgcagctaa cgcattaagc atcccgcctg gggagtacgg tcgcaagatt aaaactcaaa 780
ggaattgacg ggggcccgcg caagcgggtg agcatgtggg tcaattcgaa gcaacgcgca 840
gaaccttacc agcccttgac atgtcccgtg tgagtaccag agatggaact cttcagttcg 900
gctggcgggg acacaggtgc tgcatggctg tcgtcagctc gtgtcgtgag atgttgggtt 960
aagtcccgcg acgagcgcaa ccctcgccct tagttgccat catttagttg ggcactctaa 1020
ggggactgcc ggtgataagc cgcgaggaag gtggggatga cgtcaagtcc tcatggccct 1080
tacgggctgg gctacacacg tgctacaatg gcggtgacag tgggatgcag aggggtaacc 1140
ccgagcaaat ctcaaaaagc cgtctcagtt cggattgtgc tctgcaactc gagcacatga 1200
agttggaatc gctagtaatc gcagatcagc acg 1233

```

<210> 99

<211> 1304

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 99

```

cgaaatcccg cagggatcag tagagtggca aacgggtgag taacacgtgg gtgacctgcc 60
ttcgagtggg ggataacgtc ccgaaagggg cgctaatacc gcatgacatc ctgctcttga 120
acgagtggag atcaaagctg gggatcgcaa gacctagcgc tcaaagaggg gcccgcgcct 180
gattagctag ttggtggggg aacggctcac caaggcgacg atcagtatcc ggcctgagag 240

```

```

ggcggacgga cacactggga ctgagacacg gccagactc ctacgggagg cagcagtggg 300
gaattgttcg caatgggcgc aagcctgacg acgcaacgcc gcgtggagga tgaagatcct 360
cgggtcgtaa actcctttcg atcgagacga acggcctccg ggtgaacaat ccggaggagt 420
gacggtaccg agagaagaag ccccggtctaa ctccgtgccg gcagccgcgg taatacgggg 480
ggggcaagcg ttgttcggaa ttactgggcg taaagggctc gtaggcggcc aactaagtca 540
gacgtgaaat ccctcggctt aaccggggaa ctgcgtctga tactggatgg ctagaggttg 600
ggagagggat gcggaattcc aggtgtagcg gtgaaatgcg tagatatctg gaggaacacc 660
ggtggcgaag gcggcatcct ggaccaattc tgacgctgag gagcgaaagc caggggagca 720
aacgggatta gataccccg tagtcctggc cctaaacgat gaatgcttgg tgtggcgggt 780
atcgatccct gccgtgccga agctaacgca ttaagcattc cgcctgggga gtacggtcgc 840
aaggctgaaa ctcaaaggaa ttgacggggg ccgcacaaag cggtaggagca tgtggttcaa 900
ttcgacgcaa cgcgaagaac cttaccagg cttgaacagc gagtgaccac tcctgaaaag 960
gagcttccgc aaggacactc gtagaggtgc tgcattggctg tcgtcagctc gtgtcgtgag 1020
atgttgggtt aagtcgccga acgagcgcaa cccttgtttg ctgttgccat caggttatgg 1080
tgggcactct gcaaagactg ccggtgataa accggaggaa ggtggggatg acgtcaagtc 1140
agcatggcct ttatgtctgg ggctacacac gtgctacaat ggccggtaca aaccgtcgca 1200
aaaccgtaag gtcgagctaa tcggagaaag ccggtctcag ttcggatcgt cggctgcaac 1260
tcgccggcgt gaagttggaa tcgctagtaa tcgcggatca gcac 1304

```

<210> 100

<211> 1197

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 100

```

tctagtggcg cacgggtgcg taacgcgtgg gaatctgccc ttgggttcgg gataacagtt 60
ggaaacgact gctaataccg gatgatgtct tcggaccaa gatttatcgc ccagggatga 120
gccgcgctcg gattagctag ttggtgaggt aaaggctcac caaggcgacg atccgtagct 180
ggtctgagag gatgatcagc cacactggga ctgagacacg gccagactc ctacgggagg 240
cagcagtggg gaatattgga caatgggcga aagcctgatc cagcaatgcc gcgtgagtga 300
tgaaggcctt agggttgtaa agctcttttg cccgggatga taatgacagt accgggagaa 360
taagcccccg ctaactccgt gccagcagcc gcgtaatac ggagggggct agcgttgctc 420
ggaattactg ggcgtaaagc gcacgtaggc ggctttgtaa gttagagggtg aaagcccgga 480
gctcaactcc ggaactgcct ttaagactgc atcgcttgaa cgtcggagag gtaagtggaa 540
ttccgagtgt agaggtgaaa ttcgtagata ttcggaagaa caccagtggc gaaggcgact 600
tactggacga ctgttgacgc tgaggtgcca aagcgtgggg agcaaacagg attagatacc 660
ctggtagtcc acgccgtaa cgatgatgac tagctgtcgg ggctcatgga gtttcgggtg 720
cgcagctaac gcgttaagtc atccgcctgg ggagtacggc cgcaaggtta aaactcaaag 780
aaattgacgg gggcctgcac aagcggtgga gcatgtggtt taattcgaag caacgcgcag 840
aaccttacca gcgttgaca tggtaggacg gtttcagag atggattcct tcccttacgg 900
gacctacaca cagggtgctg atggctgtcg tcagctcgtg tcgtgagatg ttgggttaag 960
tcccgaacg agcgcaaccc tcgtctttag ttgctaccat ttagttgggc actctaaaga 1020
aactgccggg gataagccgg aggaagggtg ggatgacgtc aagtcctcat ggcccttacg 1080
cgctgggcta cacacgtgct acaatggcgg tgacagtggg cagcaaacctc gcgagagtga 1140
gcaaatcccc aaaaaccgtc tcagtccgga ttgttctctg caactcgaga gcatgaa 1197

```


<210> 101
 <211> 1352
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 101
 cgacggccag tgaattgtaa tacgactcac tatagggcga attgggccct ctagatgcat 60
 gctcgagcgg ccgccagtgt gatggatatt tgcagaattc gcccttcagg cctaacacat 120
 gcaagtcgca cgagaaaagg ctccggcccc ggtacagtgg cgcacgggtg agtaacacgt 180
 aggcaatctc ccctcgagtgt gtggataacc ttccgaaagg agggctaata cagcatgaga 240
 ccacgagctc gcagagcttg tggccaaagc ggacctcttc ttgaaagttc gcgcttgagg 300
 atgagcctgc ggcccatcag ctagttggta gggtaatggc ctaccaaggc taagacgggt 360
 agctggtctg agaggacgga cagccacact ggaactgaga cacggtccag actcctacgg 420
 gaggcagcag tggggaaatct tgcgcaatgg acgaaagtct gacgcagcga cgccgcgtga 480
 gcgatgaagg ccttcgggtt gtaaagctct gtggggagag acgaataagg tgcagctaat 540
 acctgcatcg atgacggtat ctcccttagca agcaccggct aactctgtgc cagcagccgc 600
 ggtaagacag aggggtgcaaa cgttgttcgg aattactggg cgtaaagcgc gtgtaggcgg 660
 ctgtgtaagt cgggcggtgaa atcccatggc tcaaccatgg aagtgcaccc gaaactgcgt 720
 agctagagtc ctggagagga aggtggaatg cttggtgtag aggtgaaatt cgtagatatt 780
 aagcggaaaca ccggtggcga agcggccttc tggacagtga ctgacgctga gacgcgaaag 840
 cgtggggagc aaacaggatt agataccctg gtagtccacg ccgtaaacga tgaatgctag 900
 acgtgggggt gcatgcactt cgggtgtcgc gctaacgcat taagcattcc gcctggggag 960
 tacggccgca aggttaaaac tcaaaggaat tgacgggggc ccgcacaagc ggtggagcat 1020
 gtggtttaat tcgaagcaac gcgcaaacct taccacacct tgacatgtcc attgccggtc 1080
 cgagagattg gaccttcagt tcggctggat ggaacacagg tgctgcatgg ctgtcgtcag 1140
 ctctgtcgtg gagatgttgg gttaagctcc gcaacgagcg caaccctac cgccagttgc 1200
 catattcag ttgggcactc tgggtggaact gccggtgaca agccggagga agcggggatg 1260
 acgtcaagtc ctcatggccc ttatgggttg ggctacacac gtgctacaat ggcgggtgaca 1320
 gtgggacgcg aagtccaaga tggacaaatc cc 1352

<210> 102
 <211> 1361
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 102
 aacagctatg accatgatta cgccaagctt ggtaccgagc tcggatccac tagtaacggc 60
 cgccagtgtg ctggaattcg cccttcaggc ctaacacatg caagtcgaac ggatccttcg 120
 ggattagtgg cggacgggtg agtaacacgt gggaaacgtgc cctttgggtc ggaacaactc 180
 agggaaactt gagctaatac cggataagcc ttctgagggg aagattttat gccattggag 240
 cggcccgctg aggttagct agttggtgag gtaaaagctc accaaggcga cgatccttag 300
 ctggtctgag aggtatgatca gccacattgg gactgagaca cggcccaaac tcctacggga 360

```

ggcagcagtg ggggaatcttg cgcaatgggc gcaagcctga tccagccatg ccgcgtgagt 420
gatgaaggcc ttaggggttg aaagctcttt caccggagaa gataatgacg gtatccggag 480
aagaagcccc ggctaacttc gtgccagcag ccgcggtaat acgaaggggg ctagcgttgt 540
tcggaattac tgggcgtaaa gcgcacgtag gcggatatatt aagtcagggg tgaaatccca 600
gagctcaact ctggaactgc ctttgatact gggatatcttg agtatggaag aggttaagtgg 660
aattccgagt gtagaggtga aattcgtaga tattcggagg aacaccagtg gcgaaggcgg 720
cttactggtc cattactgac gctgaggtgc gaaagcgtgg ggagcaaaca ggattagata 780
ccctggtagt ccacgccgta aacgatgaat gttagccgtc gggcagtata ctgttcgggtg 840
gcgcagctaa cgcattaaac attccgcctg gggagtacgg tcgcaagatt aaaactcaaa 900
ggaattgacg gggggcccgc caagcgggtg agcatgtggt ttaattcgaa gcaacgcgca 960
gaaccttacc agctcttgac attcgggggt tgggcagtgg agacattgtc cttcagttag 1020
gctggcccca gaacaggtgc tgcattggtg tcgtcagctc gtgtcgtgag atgttggggt 1080
aagtcocgca acgagcgcaa ccctcgccct tagttgccag catttagttg ggcactctaa 1140
ggggactgcc ggtgataagc cgagaggaag gtggggatga cgtcaagtcc tcatggccct 1200
tacgggctgg gctacacacg tgctacaatg gtggtgacag tgggcagcga gacagcgatg 1260
tcgagctaatt ctccaaaagc catctcagtt cggattgcat ctgcaactcg agtgcataaa 1320
gttgggaatcg ctagtaatcg cagatcagca tgctgcgggtg a 1361

```

<210> 103

<211> 1300

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 103

```

catgtttagt agcaatacta aatgatgacg agcggcggac ggggtaggaa cacgtaggaa 60
cctgcccagg agagggggac aaccaaggga aactttggct aataccgcat aatctctacg 120
gagaaaagtt gcccgtaagg gtggcgcttt tggaggggcc tgcgtccgat tagttagttg 180
gtgaggtaat agctcaccaa gactgtgatc ggtaactggt ctgagaggac gaccagtcac 240
actgggactg agacacggcc cagactccta cgggaggcag cagtggggaa tcttgacaaa 300
tgggggcaac cctgatccag cgatgcccg tgggtgaaga aggccttcgg gttgtaaaagc 360
cctttaggcg gggaagaagg atatgggatg aataagcctg tattttgacg gtaccgcgag 420
aataagcacc ggcaaactct gtgccagcag ccgcggtaat acagaggggt cgagcgtaa 480
tcggatttac tgggcgtaaa gggcgcgtag gcggttgtgt gagtgtgatg tgaaagcccc 540
gggctcaacc tgggaagtgc atcgcaaacg acacaactgg agtatatgag aggggtggcgg 600
aatttcgggt gtagcgtgta aatgcgtaga gatcggaagg aacgtcgatg gcgaaggcag 660
ccacctggca taatactgac gctgaggcgc gaaagcgtgg ggagcgaaca ggattagata 720
ccctggtagt ccacgccgta aacgatgaga actagatgtt ggagggggaa cccttcagta 780
tcgaagctaa cgcgataagt tctccgcctg ggaagtacag tcgcaagact gaaactcaaa 840
agaattgacg gggggcccgc caagcgggtg agcatgtggt ttaattcgat gcaacgcgaa 900
gaaccttacc tacccttgac atcctgcgaa tcttgccgag aggtgagagt gccgcaagg 960
gcgcagagac aggtgctgca tggctgtcgt cagctcgtgt tgtgagatgt tgggttaagt 1020
cccgtaacga gcgcaaccct tgccttagt tgccatcatt tagttgggga ctctaaggag 1080
accgcccgtg atgaaccgga ggaaggcggg gacgacgtca agtcatcatg gcctttatgg 1140
gtagggctac acacgtgcta caatggggcg tacagagggg cgccaacccg cgagggggag 1200
ccaatctctt aaagcgtctc gtagtccgga ttggagtctg caactcgact ccatgaagtc 1260
ggaatcgcta gtaatcgcg atcagcagtg ccgcggtgaa 1300

```

<210> 104
 <211> 1250
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 104
 tgtagcaata catcagtggc agacgggtga gtaacacgtg ggaaccttcc tcgttgtacg 60
 ggacaactca gggaaacttg agctaatacc gtatacgtcc gagaggagaa agatttatcg 120
 caatgagacg ggcccgcgtc ggattagcta gttggtaagg taacggctta ccaaggcgac 180
 gatccgtagc tgatctgaga ggatgatcag ccacactggg actgagacac ggcccagact 240
 cctacgggag gcagcagtgg ggaatcttgg acaatgggag caagcctgat ccagccatgc 300
 cgctgagtg aagaaggcct tagggttgta aagctctttt gccagggagc ataatgacgg 360
 tacctgagaa taagccccgg caaacttcgt gccagcagcc gcggtataac gaagggggct 420
 agcgttggtc ggatttactg ggcgtaaagc gcacgtaggc gggtcgttaa gtcaggggtg 480
 aaatccccga gctcaactcc ggaactgcct ttgatactgg cgaccttgag gctggaagag 540
 gttagtggaa ttcccagtg agaggtgaaa ttcgtagata ttgggaagaa caccagtggc 600
 gaaggcggct aactgggtcca gatctgacgc tgaggtgcga aagcgtgggg agcaaacagg 660
 attagatacc ctggtagtcc acgcgtaaa ctatgggtgc tagctgtcag cgggcttgct 720
 cgttgggtggc gcagctaacg cattaagcac cccgcctggg gagtacggtc gcaagattaa 780
 aacttaaaag aattgacggg ggcccgacac agcggtgag catgtgggtt aattcgaagc 840
 aacgcgcaga accttaccaa cccttgacat cccgatcgcg gacaccagag atggagtcct 900
 tcagttcggc tggatcggag acaggtgctg catggctgtc gtcagctcgt gtcgtgagat 960
 gttgggttaa gtcccgaac gagcgcaacc ctgccttta gttgccatca tttagttggg 1020
 cactctaaag ggactgccgg tgataagccg gaggaagggtg gggatgacgt caagtccca 1080
 tggcccttac gggttgggct acacacgtgc tacaatggcg gtgacaatgg gcagctactt 1140
 cgcaaggaga agctaatacc aaaaagccgt ctgagttcag attgcaactct gcaactcggg 1200
 tgcataagtg tggatcgct agtaatcgct aatcagcagg tagcgggtgaa 1250

<210> 105
 <211> 1302
 <212> ADN
 <213> Organisme Inconnu

<220>
 <223> Origine de la séquence :Organisme du sol

<400> 105
 ggcttcggct ccccggtaga gtggcgagc ggtgagtaac acgtgggtaa tctgcctttg 60
 ggtggggaat aacccttcga aagaggggct aataccgcat aacgcagcgg caccgaatgg 120
 tgacagttgt taaagtgggg gatcgcaaga cctcacgcct gaagaggagc ccgcgcccga 180
 ttagctagtt ggtgcggtaa tggcgtagca aggcggcgat cggtagccgg cctgagaggg 240
 cggacggcca cactggcact gagagacggg ccagactcct acgggaggca gcagtgggga 300
 attttgggca atgggcgcaa gcctgaccca gcaacgcgc gtgaaggacg aaatccctct 360
 gggatgtaaa cttcgaaagt tggggaagaa atccgtgtga ggataatgca cacgggatga 420

```

cggtagcccaa cgtaagcccc ggctaactac gtgccagcag ccgcggtaat acgtagggggg 480
caagcggtgt tccgaattac tgggcgtaaa gggcgcgtag gcggtacgac aagtcctggag 540
tgaaagcccc gggctcaacc ccggaatgtc tttggaaact gtcgaacttg agtcggaag 600
aggcatctgg aattcccagt gtagcggtag aatgcgtaga tattgggaag aacacctgag 660
gcgaaggcgg gatgctgggc cgacactgac gctgaggcgc gaaagccagg ggagcgaacg 720
ggattagata ccccggtagt cctggcccta aacgatggat acttgggtgt tggggttctc 780
gaagtccccc cgtgccggag ctaacgcggt aagtatcccg cctggggagt acggtcgcaa 840
ggctgaaact caaaggaatt gacggggacc cgcacaagcg gtggagcatg tggttcaatt 900
cgacgcaacg cgaagaacct tacctgggtt aaatcctacc tcgtcgctc agagatgagg 960
tttcccttcg ggggaggttag gacggtgctg catggctgtc gtcagctcgt gccgtgaggt 1020
gttgggttaa gtcccgcaac gagcgcaacc cttaccacta gttgccagcg gttcggcccg 1080
gcactctatt gggactgccg gtgacaaacc ggaggaaggt ggggatgacg tcaagtcac 1140
atggccttta tgtccagggc tacacacgtg ctacaatggc cggaacaaag cgcagcaaac 1200
ccgcgagggg gagccaatcg caaaaatccg gtctcagttc ggattggagt ctgcaactcg 1260
actccatgaa gttggaatcg ctagtaatcg cggatcagca tg 1302

```

<210> 106

<211> 1281

<212> ADN

<213> Organisme Inconnu

<220>

<223> Origine de la séquence :Organisme du sol

<400> 106

```

tgcttctctt gagagcggcg gacgggtgag taatgcctag gaatctgcct ggtagtgggg 60
gataacgttc ggaacaggac gctaataacc catagctcct acgggagaaa gcaggggacc 120
ttcgggctct gcgctatcag atgagcctag gtcggattag ctagtgtgtg aggtaatggc 180
tcaccaaggc gacgatccgt aactggtctg agaggatgat cagtcacact ggaactgaga 240
cacggtccag actcctacgg gaggcagcag tggggaatat tggacaatgg gcgaaagcct 300
gatccagcca tgccgcgtgt gtgaagaagg tcttcggatt gtaaagcact ttaagttgga 360
aggaagggca gtaaatattat actttgctgt tttgacgtta ccgacagaat aagcaccggc 420
taactctgtg ccagcagccg cggtaataca gagggtgcaa gcgttaatcg gaattactgg 480
gcgtaaagcg cgcgtaggtg gtttggttaag ttggatgtga aatccccggg ctcaacctgg 540
gaactgcatt caaaactgac tgactagagt atggtagagg gtggtggaat ttcctgtgta 600
gcggtgaaat gcgtagatat aggaaggaac accagtggcg aaggcgacca cctggactaa 660
tactgacact gaggtgcgaa agcgtgggga gcaaacagga ttagataccc tggtagtcca 720
cgccgtaaac gatgtcaact agccgttgga agccttgagc ttttagtggc gcagctaacg 780
cattaagttg accgcctggg gagtacggcc gcaagggtta aactcaaata aattgagggg 840
ggcccgcaac agcggtaggag catgtggttt aattcgaagc aacgcgaaga accttaccag 900
gccttgacat ccaatgaact ttctagagat agattggtgc cttcgggaac attgagacag 960
gtgtgcatg gctgtcgtca gctcgtgtcg tgagatgttg gggttaagtcc cgtaacgagc 1020
gcaacccttg tccttagtta ccagcacgac atgggtgggca ctctaaggag actgccggtg 1080
acaaaccgga ggaaggtggg gatgacgtca agtcacatg gcccttacgg cctgggctac 1140
acacgtgcta caatggctcg tacagagggg tgccaagccg cgagggtggag ctaatccac 1200
aaaaccgatc gtagtcggga tcgcagtcgt caactcgact gcgtgaagtc ggaatcgcta 1260
gtaatcgca atcagaaatg t 1281

```

<210> 107

<211> 43

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 107

cgctgcagat ttaaataatgc aacgcgtaag tcgatggcgt tcg

43

<210> 108

<211> 51

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 108

cgggtcaactt aattaagata tctcgagaga tctattaata cgatacctgc g

51

<210> 109

<211> 29

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 109

aaaaagatat ctgacgtccc gaaggcgtg

29

<210> 110

<211> 32

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 110

aaaaaagatc tggctaacta actaaaccga ga

32

<210> 111

<211> 36

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 111

gtgccgttaa ttaagctccg cgaagtcgct cttctt 36

<210> 112

<211> 36

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 112

gtgccgttaa ttaaccgctg cataaccctg cttcgg 36

<210> 113

<211> 42717

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:cosmide
a26g1 brin non codant

<400> 113

```

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaacc gggccctgac 60
gttcagaact ccccgcgaga atctctcggc agagcgccctg cacctcgact tcaccggcag 120
tgtcgagatc gatgcagggtg caagcgaact cgggatgctc ggccgcgatg gtacgtccca 180
gacccacac cggtgcccgc gcgggcacca ccggcatctg cagatgctcc gcatgaacgc 240
ccgccgtaat cagccagatt cgcggatgct gcgattcggc gtcggacgct tgtttcgtca 300
gctgtatgcg tccactccg aattcttgaa cgatgcgcaa aatgtcttcg caggcgggttc 360
cccccgcagc cgacggatcg gtctcatcct cggctctcgc caggctcccg cagtgaatga 420
cgccccggta aggetgatcc ggtatctcgg catcgcgacc cgaaatcaca accgtgtttg 480
tgcccgccc tcgcgccacc gcggctgcga tgccgcccgc atcggaatg accagccatg 540
cacccgacac cgttgccgtt ggactctcgg ccagcagctg cggctcccat tccatagcgt 600
ggaaccattc ggattggcgc tccagttcct gggcacgcag gccttggacc tcgaggatga 660
cgtggccttc cgcacacac agggtgacat cgccctcgag ccgccccgtc agccgcgcac 720
gcaccctaag atcgccggcg ggtctgccga aacagtgcaa ccgttcgatg gcgacaggca 780
cgcaaggacc ggcgtgcct tcgccgccaa gcgtcgcgcc cagcacctgc aaacaggcat 840
cgagcaaggc aggatgaagc gtgtaaccgg actctgcttc gcgaacggca tccggcacgc 900
tcagtgcgc cactgcctcg ccgtcgcgcc gccacacttc cgcgatgccg cggaagggtg 960
cgccgtaatg catccctgc gatgcgaagg ccgcatagaa gtcacgccc tcgatgcgat 1020
ccccaagtgt gggcaggctc accgtgggcg cgacctgtc cggcgccgca gccatggtgc 1080

```

cgcgcgcggtg ctccggtccaa tcggaaccgc cttcgggccag actggagatg cggaatgcgt 1140
 gtccctcgag tatgacctgc actcgcgagg cgcccgcgga aggaacaacc agcatttgtt 1200
 caaacccgat ctcttcagg ctgcagccac ccgcgaacac ttcttgggt gcggccagcg 1260
 ccatctccac ataagcggca ccgggaagca ccacaagctc gttgagccgg tgatcggcga 1320
 gaaacggcag cgcattccaga gagagcacgg actcccagac gtgtgtgtcc ggcggccagcg 1380
 caatctcgac cttgcgaccg agcagcggat gaccgccaac tgccggcaaa cttcgcccgcg 1440
 tcgaggtcgc gaaccagaag cgctcgcgct gccagggata cgtcggcaga tccaggcgcg 1500
 tgtcgggaga cgaagcgagc gcgcgccagt ccggacgctg cccattcacg tagagcgcg 1560
 cgagcaactc gagcagctca cgccgctccg gttcgtcgcg gcgcagtacg gggcgaacca 1620
 gtccgtttat gccgagcgtc cgcagactat cctcgatcga cggcgctcagc acaggatgcg 1680
 gactgatctc caggaactgc gtgaactcat caccagccat cgctgcaac gactcccaga 1740
 aacggactgg ctgtcgcaga ttggctaccc agtacgacgc gtgcacgcc tcgcccggtc 1800
 tcaactgtcc ttcaaccgtg gagaagaacg gcacggcgga acgttttgca ataacgcggc 1860
 cgagttcctg gcgcaattcg ttctcgagcg ggtccacctg cgagctgtgt gaagcgacat 1920
 ccacctgaat cagccggcag aagacgcgcg gcctctcgaa gtctccttc aaatgctcga 1980
 gagccacacg gtctcccgag aacaccgtgc tgcgtggctc gttgctggcc gcgacagaaa 2040
 cagtagtgag acccggttca gcgagcacgg ccttcgcccg atcgagcggc agttcgacca 2100
 gagccatcgc tccccggccg cgaagtccga gcaacagccg gctgcggcga cagatgatgc 2160
 gggccgcgtc ctccagggtg agaatgctg cgacatgggc cgccgccact tctcccatgc 2220
 tgtgtccggc cagccggtcc gggcgaattc cccaggattg cagcagttcg accagcgcg 2280
 tttgaacggc gaacagcgca ggctgcacgc gatcgatctg gctcagccat gctcccgact 2340
 cgtcggcgag caggtccgca agccgccatt ccacgaagct gcggaaggcg gcgtcgcaac 2400
 gttcgatcgc cgatcggaag acaggctcgt cggaatacag gcgatacgcc atgcgcgggt 2460
 actgtccgcc ctggccggaa aagatgaagg cgagtttcgg acgaactccg ggatcggcga 2520
 aaccggtggc gacgcgcgca ttggtttcat tgcgcggaa ggctcagc aattgattga 2580
 actcgggcag ggatgaggcc acaaacgctg cgcgatgttc gtagtgactg cgcgtcaggc 2640
 tggcgggcga acacagcgcg gagagcgag cgtaaaagcg cccatcgcg taggcgccg 2700
 cgagatcgcg cagagcctgc ggatggcgcg ccgaaagcg taggaggtat tcgcggccgt 2760
 cttcggcagc gagttgatcc gggaccggca cttctcgcg gctgggcggc cggcccgacc 2820
 ccccgccgcc gcttcgcggg cggtgtccc cctcgaggag ggggacacta gcactagcct 2880
 tagcctccac attccctttg gcctctacac tcgccttgac ctcttactc gtcttcgct 2940
 ctacattcgt cttcgctct acattcgtct tcgcttctac attcgtcttc gcttctcgca 3000
 ggacgacgtg cgaattggtg ccactgattc cgaacgagct caccggcgca actcgggggc 3060
 ggccgttga gggccatggc gaacatgcgg ttgctatctt tagcggcagc tcattccaga 3120
 gtacgtgcgg gttgggcggc ttgaaatgca gatggggcg aatctctcgg tgctgcagg 3180
 cgagaatggt cttgatcagg ccggcgatac ctgccgccgc ctccagggtg ccgaagtgg 3240
 ttttcaccga cccgacgac aacggagaat cgacggcacg cccctcgccc agcaccgctg 3300
 ccatcgccc cagttcgatg ggatctccca gcggcgctcc ggttcggtg gcttccacgt 3360
 aatcgacatc ggcgggggccc atgccggcgt tcttgagcgc cgcccgaaac acggcttcc 3420
 gcgcgggacc gttcggcgc gtgagggcgt tgctgcggcc gccgtggtt acggccgatc 3480
 cgcaatcag cgccagaata cgatcgccgt cagcgtcgc atcgacagc cgcttcagca 3540
 ccagcattcc gcatccctcg ccgcggcgt aaccgtcggc ggaggcagcg aaacttttgc 3600
 aacggccatc ggccgccatg gcccgaggc ggcagaagta gatcgtgct tccggcgcca 3660
 gaatcagggt cagccgcgc gccagcgcca tgctgcaact tcgcgactgc aagctgcggc 3720
 acgccagatg aaccgccacg agtgaggga agcacgccgt gtcgacgggg aagttcggtc 3780
 cctgcaaccc cagcagatag gagatccgtc cggcggcagt gctgaacgcg gttccggtac 3840
 cggatatagg gtcaatgagc gccggatcgg taggtttcag ccggctgtag tcgtcgggtc 3900
 tgatcccgat gaacactccg gtgtcgtcgc ccgcgagact gtcggggcgcc cgaccgcac 3960
 gctccaaagc ttcccatgcc acctcgagca gcaggcgctg ctgcggatcc agaccggcga 4020
 cctcgcgcg cggtgattcc aagaagccgg cgctcgaagc gtcgacggca ccatcgagga 4080

atccgcccag acgcggtgtac atctttcccg gcgcgttggg atcgggatcg taaaacgcat 4140
 cggcatccca acgcccgcga ggaatttcgc ggatggcatc gatgccatcg tgcaggagct 4200
 gccaaaatgc ttccggcgag tccgcgcgg gaaagcggca agccatgccg acgatcgcg 4260
 tgggttcgtt gtggacgctc tccagttcgt cgagacgcgc tcgctgcgc ttgagcgcca 4320
 ggaccgcctg ttgaagagga gtgagatcgc tcatcgttcc tctcccagat ggtgcaaggt 4380
 ttccgtgatg aacgcccga tctcctgctc ggacagctgc cgcacttcat cgacgaggga 4440
 atcgctgggg acgtcgagtc cgagttcgcg gaggacatgg ccggccaatt tctcgacggt 4500
 cggatggctg tatagcaatg tcgcgggaag gctcttgcc accagctctc cgatggcgcg 4560
 cgccagatcc agcgccatca gcgaatcgag tccgtattcc ttgagcgcc ggcgcggtc 4620
 gagcgtcttg gacgcacga gcgccagcac gccgcggcc tgcttgcgga tgcgcactc 4680
 cagcagttcc tgccgcgcct ccggcgcgag ttcggtgagt tgctggatga agccgggac 4740
 gctcgccggg tccgccttt ttccggcgga gacttggaa acccgcatc gagcggcagt 4800
 ctcgcccagc agatgcgcga agatgcgcgc acccacttcc ggcgcgagca gcggtacccc 4860
 cggcaggcct tgccgcgcga tgcgcgcgcc catgccttcg ccgcccattg gtccccatt 4920
 gatgctcagc gccggtagtc cttgcgcgcg gcgcagtgg gcaaggctgt cgagaaatgc 4980
 gttggccgcc gactaatgc tctgtccggc ggaaccgagc agcgaagcgg cggaagagaa 5040
 gactacgaaa aagtcgagcg catggtggcg agtgagctgg tgaagattcc aggcacctg 5100
 cagcttcggc gccagcacct tctcgaaacg agccacgtc tgttctgtaa ctacccctc 5160
 atcgagcacg cctgcggcat gcacgactcc acgcagcggc tgggtgcgcg ggtccgccag 5220
 cagcgccgcg agctgttgct cggaactcac atcgcaagca gcaaccatga ctgcagcccc 5280
 gagttgctcg agatcggaac ctgcctccgt atgcggccg accagtacca gacggcgcg 5340
 gccttgctcg atcaagcggc gtgccacct tcgtcctaat gcgcccagac cgccggtgat 5400
 cagatagacg ccgtcggtc aaatggcagg cggccgcttc gacgtttcct tgtgccgcac 5460
 cagccggcga acgtagcggc gtccgttgcg caatgcgac gctttgtcgt cgccggcata 5520
 acggatttca tccagcagca tggcgcgcg gatgtcgga ttgtcgcaac cgaggtcgat 5580
 caggccgccc cacagctcgg gatgctcgcg cgcgatcgcc tgcccagtc cccacagtgg 5640
 agcctggaaa ggatcgacgg gactcgcatc gtcactactg atgcgatgca cgccgcgcgt 5700
 gatcagccag agccgcgcgg ggcgccgcac caaagtctgg gtctgttcca gcgcctggcg 5760
 gagagtacg ccggacacga cgcgcagctc ttgcggcatc aaaccggcga catcgtctgc 5820
 gccggcacac agcaaccagt cctccggctt gccaggggcg tcggacttca acgttgtcga 5880
 tcgctcgcac tctgcccact gcacatcctg cagccacgac tgtgcggatt gcagcgtacc 5940
 ggcatgcatt acagccagtc cgaaaaactc ggcgatgacc gcgcccgtct cttcaaccag 6000
 cgtcagatca ccgacgaacg ggccgctcga gctcgggcgc agacgcgcac gacagcgag 6060
 agaacctgcc ggcgacgggt agaagcgcac cgcttcgac ccgaccggca cgtatgcgcc 6120
 gggctggcaa cgctccgcgg gccaaagtcg tccgaatact tgaacaag aatcgatcag 6180
 gccggggtgc agccggtaag cgttcgcgcc atcctcagcc accggcagac gcattcgccc 6240
 cagcgctcgc ccacgcgcac gccagacttc ttccacccaa ctgaaggcgg ggccaagatc 6300
 gacgccgcgt gcgttcacgc cgcgtagaa cgcacgcgc gaaatgactt cggaaggctg 6360
 cgccggcagc tcgaaatgaa cggcgccggc agtcgcgcgc cgcagactgg ctgccgtgtg 6420
 gagcttccac gaatcgccat cctggctgaa gacctgcacc tttgcttcgc cgtcctcgcc 6480
 ggggtgtgaca atcgcttgca ccgtgaccgg cgtatccggc gggatggcca gtgcctgccg 6540
 catcatgaca tcggagacgg cgcagggaac cggaccgaag acttctctgt ccgcttcgag 6600
 aaatgccgac acgtgcccag cgcggggcac aatgaccgc tcgtagatca cgtgctcatg 6660
 gagcagaggc gtctccgtgg ttagcgaatt ttgaagatg acatcgccca acgcgtgtt 6720
 gaggcgcgct cccaacatgc cgcgcgcgc cggctctctc gcgggtacgc gtctcaggct 6780
 gaagggtgtc cgctgaaacg gatacgtcgg cagcgcgac cggctgggtg attccccggc 6840
 atagagaccg cgccagtcgg gattcacgcc cgcggtaaac aggcgcgcaa gactttccag 6900
 cagcacggac caatccgatc gtcccttaga tagggagtgc agccagaccg cgccgtcatc 6960
 gggcagacaa tatcgcccca gcgtgggtgag cgtgggatgc gggccgattt ccagaaacag 7020
 cttgcaactc cggctccgca gggttcgcat cgcgctttca aactgcacgg tttcgcgcaa 7080

ctgtcgccgc	cagtagcggg	cgtcgagtg	cgctgcctt	ggcaatacgg	ctccgctgac	7140
gttcgacacc	agcgggatcg	ccagcggctg	atacgcgac	gcacctgcaa	gcgcttcgaa	7200
cttgtccaaa	atcggatcca	tcagcggcga	atggaacgca	tgcgatacgt	tcagctctcg	7260
cgtttccacg	ccggcgcgat	gcaggtcatc	ttgcgcttcc	gcgatttctg	cagccgtgcc	7320
ggagatcacg	gtgcggtccg	gcgcattcga	tgcggcgact	gccaccttgg	cggcgagcgc	7380
cgcgatgcgg	ctcggattgg	cgtgaacgat	gaccgctttg	ccgcggggaa	gcgcattgac	7440
cagccgcccc	ctggcggcca	ccagccgcag	gccgtctctc	acgctaaagg	cgcgggccac	7500
acacgccgca	acatactcgc	cgagactgtg	gcccagcacg	tagtccggcc	ggacgccgag	7560
cgacagccag	aactgagcca	gagcccatc	cagggcaaac	attgccggct	gggtatacgc	7620
cgtctcgtgc	aacggcgaaa	ccgactcgaa	caagagaacg	gtcagcggaa	catcgagctg	7680
gggacggagc	caatcggcgc	aacgatcgag	cgcgtcgcca	aaaacaggct	gcgttttata	7740
aagctctgcy	cccattgccgg	cgtactgcgc	gccctggccg	gtgaagagaa	aagcgattgc	7800
cggccgcccc	cgcaacgata	cttcgcgcgc	cggcgccgcg	gccaatgccg	ctacagcctc	7860
tgcgcacatc	gcggcgggtga	tcgccaagcg	gtgactatat	gcgtcgcgcc	caacctgact	7920
ggtgaagcaa	acgtcggaca	gcaacgcatt	cgggtgcgac	tgaggaact	ccgcgaagtg	7980
gccggccagt	tcgcccagcg	cttcgtcgg	gcgcgccgac	agagtgcgaa	gctgcggggc	8040
tgtgaccggc	ttcggcaaa	ggagtgcagg	cgctcttcg	aggatgacgt	gcgcgttgct	8100
ccctccaaaa	ccgaacgagc	tgacgccggc	cagacgcggc	cgtccttcg	acgtccacgg	8160
cgacgattcc	gtggcgtatc	gaaaccggct	gccgtccagt	gagatgttcg	gattcagccg	8220
gcgaaaatgc	aggtgcggag	gaatggtgcg	atgctgcagg	gcgagtacgg	ctttgatcag	8280
cccggcgatt	cccgcgcgc	cctccagatg	cccgatgttg	gtctttacgg	aaccagcag	8340
acaaggcgca	gagtcggcg	cgctcgtagc	cgactgcagg	gcctcgatct	cgataggatc	8400
gcccagcgac	gtgcccgtgc	catgcgcctc	gatcaacgat	acgtgggatg	gatcgatgtg	8460
cgcgttgggc	accgcctctt	gcaggaccgc	cttctgcgcc	tgagattcgc	gcgcgctgat	8520
gccattgtct	cgtccgtcct	gattgattgc	cgagccgcgg	atgactgcac	ggatggcatc	8580
gccatcgccc	agcgcacccg	agagccgctt	cagcagcacg	atgccgcagc	cctcgccgcg	8640
cacataaccg	tcggctgcgg	cgtcgaaagt	cttgacagct	ccgtcggggc	ccaacatgcg	8700
agccttcgac	aaagcgatca	tgccctcggg	agtcaggatc	aagttcactc	cgcggcgcaa	8760
tgccgcacgc	cattcgcgcc	ggcgaggctt	ttggcaagcc	agatggacgg	cgacgagcgc	8820
ggaggagcag	gccgtatcga	ccgcatgctt	cggaccgcgc	aggtcgagca	gataggagat	8880
gcgattggcc	aacatgctat	gcgccacgcc	ggaacccgac	caagctccga	tgccggcgagg	8940
gtcggcgatc	tgaaacagtc	cgaagtcctg	ggcgagggag	ccggcaaaag	cgcgggtcgc	9000
gctgcccggc	agagggccgg	gagagatgcc	ggcgctctct	gccgcttccc	agcacacttc	9060
cagcagcagc	cgtgctgcy	gatccatgtt	cagagcttcg	cggggggaga	tgccgaagaa	9120
ttccgcacgc	aaaccgtcaa	tgcttcgag	gaaggcggca	tatcgcgcat	acgccttgcc	9180
cggagcatcg	ggatcggagg	agtagtactg	gtccgagttc	cagcggtctg	gcggcacctc	9240
ggtgacaccg	tcgacaccgt	tcttcaacag	cgtccagaag	gcgtccggat	tcttcgcgcc	9300
gcccggaaaa	cgacacgcca	tgccgacgat	ggcgatgggt	tcggcgtaga	ccaggtcgaa	9360
gctcgcgatg	ttttgccgca	tgttccgcgc	caatagcgcg	agtttgaccg	acgacatggg	9420
cgcaattttt	tccctcacga	ccttgctcct	cggagcgcag	ccacggctgc	ttcttccgac	9480
atgtcgtcca	acccgctgag	tgagtttttc	atggcgcggc	tgtctccttc	cgcagcagcc	9540
gcggcctgcy	cttcgaccag	cggcagtcct	atttgcgacg	ccaggtgcgg	ggcaagaccg	9600
gccagcgtgg	gatgacccca	aatcagggtc	gcggggagcg	tgagacccag	tgtagattcg	9660
agacggttgc	gaaactccag	ggccatgagg	gaatcgaaag	cgagttcctt	cagcggggcg	9720
aggggatcga	tagtttgaga	gtcgatgcgc	agcacgcgcg	ccagctgctg	ctgtagatgt	9780
tcttcgagca	atgtcctgcy	ggtctgaggc	tcggccgatt	gcagccgcgc	gcgcaacgcg	9840
tttggcgcac	cggcttcgct	cgcgcgctcg	tcatgcaaaa	gctcgaaacg	tgagactgc	9900
gccgccttgg	gatagaactg	ccgccactgg	cggacattga	tgggcatcgc	ggcgacgtgg	9960
caagccgagc	tgttcagcag	ctgttcagga	atagcgaggc	cgtgttgccg	cgtcaggttt	10020
tccatgccgc	gcaaagccag	ccgcgatccg	cgattgtcct	gcgcggcgag	cagccccgac	10080

tccgaccacg caccccaacc gatgctcagc gccggcaggg cttgggcctt ccggtagtag 10140
 gccagcgcgt caagaaaaggc gttcgcggcc gcgtagtttc cctgggcggg cgcgccagc 10200
 agtcctgcag cggaggagaa gagcacgaaa tgatcgagcg ggcagtcgcg ggtgagcaag 10260
 tgcaggttcc aggcaccgtc gatttttcgcg gccatcacgt tgcggaaatg cgcttccgtc 10320
 tggttcagta gcagcgcacg gtcgagaacg gctgcggcat gaatcacgcc gcgcaatcga 10380
 tcgatggaag agatcacgcg ctcgagttca tgcgctgag aaacatcgcc ctgcaccgtc 10440
 cggacatctg cgtccatgac ggcgatggct tgctggacct cgggtgaagg cgcgcggcgg 10500
 ctacgcagca ccagccggcg ggcgcggcgt ccgatcatcc agcgtgcgac ggtaagaccg 10560
 agcccgccaa gtccgcgggt aatcaagtag gttccctcgc tatcgaacgc cgagcgtagg 10620
 ggtgcgatgg gcgcattggc gcaatctcgc atcgccatga cgattttgcc gatgtgccgc 10680
 gcctgcgcca tgggtgcgaaa cgcctccacc gattcgggtga tggctgctac tcgctttcc 10740
 agggggccgc aggtttccga ttcgaatttt gcgaccatct cctgcagcag ctcccggtc 10800
 aatgccgggc gcttcaggga catgccgagc aaatcgacca gcgtgtacga gaggttcttc 10860
 aggaacgggc gaagccccag cttgcggccg gcatagtaat cgcgcttgcc gatctcgatg 10920
 aaccgtccat gatcgcgagc cagatcgaaag ctgcctcca gcagatcgcc ggaaagcgaa 10980
 ttcaggacga cgtctactcc ttcttgatcc gtccaattgc ggatgtcgtc caccgaaagcc 11040
 atcgagcgcg aatccgaaac atgcgcgatg cccagcgagc gcagatacgc tcgtttttcc 11100
 ggactcccg cagtagcgaa gatctccgcg cccgcacgct gtgcgatctg gattgccgcc 11160
 aatcccacac cgcgggtggc agcgtgaatc aggactcgtt cgcggggcgc cagccgcgcc 11220
 gctcgcgaga gcgcgtaatc ggcgggtgaga aacgcgatag gcagggcggc ggctgttcg 11280
 gcgggaatgt tggccggctt caaggcaacg cgggaaggcg gcgtggtgac gaagcgaccg 11340
 aaactgcaag gcgcaagggc cagacttca tctccgatgc gaaagtcggt gacgccttcc 11400
 cccatggcca cgatacggcc cgagcattcg ccgccaggc gcgggctgcc ggcaatcgcg 11460
 ccgggcgcac cgtcgggcat aacgcgagg gcgagcagaa cgtcgaggaa gttcaggccc 11520
 gcggcgagca cttcaatctc cacttcaccg gcttgcgggg ggcggcgcgga tgtggccgc 11580
 aagcgagcc ggtcgaggac tccgggggca tcgatctcga gccggaacgg ccgatcgccg 11640
 gccttgaaca tggcggttg catatccgct tcgtgccgag ccacgcgcgc gacgtaacgc 11700
 gcgcgcgcgc gaaaggcgat ttgattctcg ccgttggtcg tcagcagttc gtgcaggagt 11760
 tcctcttcgc cgcggcggg atcgagatcg atcagcgtgc agttcagttc cggatgttcg 11820
 taatgcacgg tccggcccaa accccagaaa ggcgcctgag cgataccggc ttgcaggatc 11880
 tgtccatcga ccggctgcgc gccgcgcgtg accagccata ggcgcgggtc ttgacgccag 11940
 ggcgtgcgcc ccagggtctg gaggagatgc agaatgcggt cgcagagggg ttcgtgctcg 12000
 agcaaaaaca cgatttcctc gagcggcggc tggagttcat cgagcttttc cggcgaggtc 12060
 tgcgtcacgc ggttgccggt agcgcgcagc catgcggtga gcgcgctatc cacagcgccg 12120
 acaatgagcc atgaccgcgc cgtcgcgcgc gccggcggtc ctgcagcggc gtgcggctga 12180
 gcgaccagc gcagttcgtg caaccagccg cgcagtgcga tgcgtccga cgcattccag 12240
 cgctgcagcc gcagaccctc gatgcggggc accagttgtc cctctccgtc cagcagcgac 12300
 agatcggcga taggtccttc cagccgcgca tgcgtccaca ccacggaacg tgcgggatgc 12360
 agccagcgca tccggtcgat gccggcgggc agccaggttc caccggcggg accaaacgcc 12420
 gcggcgatga tctgcagaca tgcacgagg aacgcggcg cagtggaacg cgtttccgag 12480
 ctacgcagac gcccgatcgc ctacactgga caactccaga tctgctcgag cgcgcggaaa 12540
 gccggaccat actcgacgcc gtgctccgcc atctgacgcc acagctccgc cgcgggcacc 12600
 actgtggggc agcgggcctg caccgtctcc gcagaatccg gcgggacggt cgatgcatcc 12660
 gcaggcgtct gacgaatgtc cccggaagca tgcaggacct atgtcgatgc ctgccggctg 12720
 gaaatccgaa acgacgccat cccgggtcta tcgaccgca tggccagctg caacgtcatg 12780
 ctgccgtcgc gcggcacaat gagcatctgt gtgaaagtca catgctccag cagcacgga 12840
 ctttcaccga aggtctcgga agttccggcc agagccatat cgagatacgc agtagccggc 12900
 aagacgactt cgcctgcac gcgatggtct gccagccaag gcacggaagc gagactgagt 12960
 tccgtctccc agaagaaagt gccgggttg gtcgaggctt cgacgcgttt tcccaacagc 13020
 ggattgcca acgtgatcgc gtgtcgcgcg ggggaagcgt cgagccagaa acgacgacgc 13080

tgccagggat accggggcag ggcacgcaa ttgccggaag ggtacacggt ccgccatgcg 13140
 acagtgtgcc cagcctcata gagggcgccc agcgacgtga gcatggaacc gcgttcgtcc 13200
 tggtcgcggc gcagagacgg aaccagcgcc gcattgccgc cgatggcggg cagcaggatg 13260
 ggatgagggc tgatctcgag aaagacatcg tgcccgtgt cggaagatg gcggatgccc 13320
 tgccagaaca gaaccggcga tcgcagattg cgagcccagt acgtgctgtc gaggtggtg 13380
 gtctccagcg tcgcgccggc caccgtggag taaaaaggta tggtcgcggg ccgcggttga 13440
 atcccgtcga gcgactgcag gagttcgtcg cacaatgggt ccacttgcgg gctatgcgcg 13500
 gcgaagtcca ctttcaccgg ccggcaagac acgcctcgcc gctccagcgt cgcgacgacc 13560
 tcggccaggg cttcgacttc accggagatg acggtggagt tgggtccgtt cgacaccgcg 13620
 ggcgatagtc gttccgtgta agtcgacagc acggcctcac attccgcgag cggcagctcc 13680
 accatcgcca tcccgcaccag gccgtgatc cggctcaaca gccggctgcg gctgcaaata 13740
 atccgcgccg catcctgcag cgtcagcgca ccgcgcacat gagcggcggc gacctctccc 13800
 atgctgtgcc cgatgacggc atccggctcg attccccagg aacgccacaa tgcggcgatg 13860
 gcgacctgca gcgcgaagag cgcaggctga atgacctga cgcggtcgag cttcgccagt 13920
 tcttctttca gcgaccagtc cacataaggc cgcattggcg cctcgcagcg ttccaacgcc 13980
 tcgcgaataa cgggttcgcg gtccatccag ctgcgcccc ttccgatcca ttgcatccc 14040
 tgtcccgaga agacgaatac cgtcttcctg cgtgggaag ggatcgtgat cccctggagc 14100
 tgccgcgcca gttcttcagc cgttctgccc gaaacagcga gccggcatcg gtgatgcgtg 14160
 cggcggactg cggccgtgta gcaaagatca cgcaggctcg gtgctgcca cgctgtcagc 14220
 aattccccgt atgcccgcg caccgcagcg agttcgtccg caccatgcgc ggacagcgga 14280
 agcacataca tcgcgtctgc aatgcccgta gttgccgcag tgctgcagt ccctgcaatg 14340
 tcgggagtggt ctgcagtgtc gggagtgtct gcagtgtcgg gagtgcctgc aatgtcggga 14400
 gtgccccag tgtccccctc cgcgaggggg acagccgccc gcgcagcggc ggccgggggt 14460
 cgggaatgga acccgctcgc agcttcgct ctaccagtcg gcgcgcctt ttcaagaacg 14520
 acatgcgcgt tcgtgccgga ccaaccaaac gcgctgacgc ccgcaaacct tcgtctcgaa 14580
 cccgcggggc acggccggac ttcttcaca atgtcgagcg acgttccctc caaccggata 14640
 ttccgggtca gctgtctcac gtgtaagtc ggcggtatcg tctcgtgact caatgcgagc 14700
 accgctttaa tcaatcccg c tatgctgccc gctccctcca ggtggccgat gttcgatttc 14760
 agggaccgga ccgcgcacac atcgccgaca ggtcgcggga ggccgacggt ttccgccagc 14820
 gcctcgatct cgatgggatc gccgagcgga gtccccgtgc catgggcttc gatgtaaccg 14880
 atctgctgcg ccgcgacgcc cgcattggcc aatgccgacc ggatgacgac ctgctgagac 14940
 acgacattgg gagcggtag cccggccgag cggccatcct gattgaccgc ggagccgcgc 15000
 accacggccc acaccgggtc tccggccgcg agtgcacgg acaggcgctt cagcaccacc 15060
 acgccgcagc cttctccgaa cacgatgccg tccgcgcgc cgtcgaaagg gcggcagcga 15120
 ccgtggggcg aggcgggttc catcttcgag gtggcgta taaactccgg cgagaagcgc 15180
 agattcactc cgccggccac ggccagcgta cactgcgcgc tgcgcaggct ctggcacgcc 15240
 agatgaaccg ccgccagcga agacgagcag gccgtgtcga gcgcgatgct gggtccttgc 15300
 aagttcagca aataggaaag tcggccggcg atcacgctat gcgcggtgcc ggtggcggtg 15360
 tacggatcga tgcgcgcgcc atcgccggtc tgcattccaga aatagtcgct gctttggctg 15420
 tggatcccga cgaagacgcc cgtgcggctg ccggagagcc cttccatcgt ctgccccgca 15480
 tcctccagtg cctcccaagc cacttccaac agcagccgct gctgcggatc aatgctgacg 15540
 gcctcgcgtg gcgaaatgcc gaaaaaatcg ttgtcgaaac catcgatgga atcgagaaat 15600
 ccggcttgaa tcttcaccgg cgtggcgggg ttcaacgatt tcaggatgcg ccggaccgac 15660
 tcctcgtccc atcgtccagg cggtaacctc cgaatagcat cgactccact gcgcaacatc 15720
 tgccagaact catcgggccc atcgccgccc ggaaaccggc agcccagacc cacgatcgcg 15780
 atgggttcgc gcgcgtcgcg ttccggccga tcgagacgct gctgcatgtg ctccagcgctc 15840
 aggtacgcct gctgcaaccg cgtaagggtg gggaatcgct cggatatcga actcactcgg 15900
 aggtcctga aaaatgagcg aacttctgtt tcaacaaagc ttcgatttct ttgtccccc 15960
 acccgcgcat ctgggttgcg acggcgtcga gatcgtctgc agcggcgggg ctccgggtcct 16020
 cgcccgcgcc ggtgccaacg gtagcaagg tagcaacggc agcaacggtc gaagggttcag 16080

cattgccggc	catgctttcc	agcggcaggc	cgagcttgtc	ggcgagatgc	tgcgccaggg	16140
cggagaatgt	cgggtaacgc	cagatcaggg	tggcagaaag	cttgacgcgc	agcccggctt	16200
ccagacggtt	gcgaaactcg	agggccatca	acgaatcgaa	tccgagatca	cccagcgctc	16260
ctctgccgtc	gagtttcgct	ggatcgaagc	gcagcacgtg	tccggcttcg	tgcatacagca	16320
gcgtttccag	ccgcgcgcgg	cgctgccgcc	cggctggaac	tgccaggagc	tcgctgcgca	16380
tgtcggccgc	cggtttggtg	tccgcggccg	cgggtgcgat	gccggccagc	agggacatcg	16440
atgcggccga	cggatagtaa	cggagccact	gcgcgatatc	gaagttcatt	acagcgacgt	16500
gcggccgaat	ctgcgtcaat	gctttgtaga	gcgcgcgcaa	tccctgttgc	ggttgaataa	16560
ccgagatgcc	gcgcgcggcc	agacggcttc	cgcggttcgc	ctgtgcggcc	aaaccaacct	16620
gtgtccacgg	tccccacgcg	atgctgacgg	cgggaagacc	ctgggcgcgg	cgcagatgag	16680
ccagcgcgtc	gagaaatgaa	ttgccggcgg	cgtagttgcc	ctggccggga	gatcccactg	16740
tgcgcgtggc	ggaagagaag	agaacaaaat	gatccagcgg	ccggccggcg	gtgagttcgt	16800
gcaggttcca	cgcgcgggct	actttcggag	ccatggcggc	ttcgaagcgt	tcggctcgtga	16860
gattgagcag	catgccgtcg	gccagcgtgc	ctgccagatg	gaacacgccc	cgaacggcgg	16920
gcatgtcgcg	atcgatgac	gcgagcgcac	ccgatagctg	ctgccggttc	gccacgtccg	16980
catggatgat	cttgacgttg	acaccttcca	gttggtggcc	aggacgctcg	ctgcgtccca	17040
gcagaacgag	atggcgcgct	ccggcggcgg	cgagccatcc	cgccacctgc	agtccgagtc	17100
cgccgagccc	gcccgtgatc	agataggttg	cgtcggcacg	gaacgccaca	tcgggtgcgg	17160
atgggatctt	gtgaagactg	aggcgcggcg	cccataccgt	gccttgccgg	atcgcaactt	17220
gacccctctg	gatattcgac	agcatcagcg	tcgcgagatg	cccgcagtcg	ttgctgtgcg	17280
catcgagatc	gacgagcgtg	cagcgcagct	cgggatgctc	ataggcaatc	gtccgcccaa	17340
ttccgtgcag	ccaggcttgt	cgaatatcaa	tatctttgtc	ggagttgaga	accgcggcag	17400
atccgcgcgt	cagcagccac	aggcgcggcg	gctcaggcca	gcccgtttgc	acgatgctgc	17460
gcaatacggg	aagcaggtcg	tcgatgcgcg	gcgagggaca	gtacacaatt	tgacggcacg	17520
gcggacccga	gcacgtatcg	gccgtgcggc	aggtttgggc	gcgcttttgc	agagtctcgg	17580
caatcgccgg	ctcgcgatg	acgagccaag	gtccgccagc	attgccggca	ttggcatcgc	17640
cgcggcgaac	cgacgcggtc	cattgcaccg	tccaggtggg	aatctccgat	tcgccgagct	17700
ggccgctatg	ggcgactctc	gactgcagcc	ccaccaattc	cgccaccacg	ctgccggtgc	17760
cggtagacgag	acggacatcc	accgtggaat	ccggccgcaa	gaccgcgtat	ccccagaccg	17820
ggccagtggg	cacttcagcg	agcgagaatc	ggtccagacc	taccggcaca	tgacatcttt	17880
tcaaactcgt	gtgatggacg	agggccgcgg	gcaactgcag	acagcagtcg	atcgtctgca	17940
tttccgtcag	cggaaatgtc	acgcgacaaa	gcacctcacc	gttgccgcgc	cagatggggc	18000
cgatggttcg	gaaggtggga	ccgaagtgat	agccgcgatc	ccacagtcgc	gaatagaagg	18060
catcggctgt	gagctccgcc	gtgcagcggg	cgcgaatcgc	atccagatcg	atggatgccg	18120
tggaaatcgc	cgcctgcagc	atgccttcgc	tgtgcagctt	ccaggaaatc	tcgcggctgt	18180
agatgcggaa	ggaagctccg	ccgcctctt	catgacggag	taccagttga	acctgcctgg	18240
cagcatcggt	ttccggcagc	gtcagcgcgc	ccgtcaatga	cacgtgttcg	acatggtgag	18300
gcccggcgcc	gagaccttgg	cgcgcagcgg	cgagcgccat	tgccaggtgc	cacgctcccg	18360
gagtcacgat	cacatcgtgc	agccggtgat	ccgcgaaatc	tttcgcctcc	acagtggact	18420
cgaactgcat	ctccggcagc	ggcgacggga	tccgcgggcc	aggcaaagcc	tgagactcga	18480
cctgcggcgg	acggatatcg	atccaataac	gctcacgctg	ccagggatag	ttgggcagcc	18540
ggcgagtttg	gccgccgttg	ggataaatac	gagaccagtc	cggagtgact	ccgttagtca	18600
gcagcgtccc	cagcgtccgg	cgcagtgcga	ggtttccgtc	ttcatcgcgc	cgaacgagg	18660
cagcggcaat	cgtgcccga	tctccgagcg	tttctgggat	cggctggacc	aacaacgggt	18720
ggggactcag	ttccagaaac	acatcatgac	caccgcgcgc	ggctgcggcg	acggccgtcg	18780
acagcatcac	gggttgggga	agattacgag	cccagtacgc	agaaaccagc	tcttcaccgc	18840
taatcgctgc	gccggtgacg	gtggagtaca	tgccaaagggc	ggccggccgc	ggctgaagcg	18900
ctccaccac	gcccggaac	gccgcgcaca	cggagtccat	cagatggctg	tgcgaggcaa	18960
tgtccacttt	cacgcgacgg	cagaagacgt	ctttcgcttc	cagttcccg	agcagttcgc	19020
ccagagctgc	gctgtcgccc	gacaggacgg	tgtcgcgcgg	gctgttgctg	gcggcaatcg	19080

agacccgatac	cgagcgccccg	gcgatggcag	cgatggcctc	gtccagcgct	aattccacga	19140
cagccatttc	tccctggccg	cgtactccgg	cgagcatccg	gctgcgagg	caaatacccc	19200
gagcggttc	atcgagagtc	agcgacactg	caatgtgcgc	tgccgcgact	tcgccccatgc	19260
tgtggccgat	cacggcgctcc	ggctcgattc	cccaatggcg	ccacagtccg	gccaaggcga	19320
ccccgactgc	gaacaggggcc	ggttgaatca	cgtcgatgcg	gtcgagcggc	ccctgcaact	19380
cttgcgtag	cgaccagtgc	acgtaaggct	gcatggcgcg	gccgcactct	tcgatggcgg	19440
cacggaacac	cggttcagaa	gccatcaggt	cgcggcccat	gccggggccac	tgcatcctt	19500
gtcccggcaa	aacgaaaacg	acttttcgct	tctggcccg	cggcacaaaa	cctgtggcgg	19560
tatcgcggtt	cgggttgccc	gccagaaaac	tgccagccc	ggccatcaag	tcttgccgct	19620
tcgtcccgg	gaatgccgcg	cgggtgttcg	atgaagtgcg	gcgagcgac	gccgtgtagc	19680
aggtgtcggc	ggggttgtcg	ttcaccacgt	cgcggtatgc	gcgcgccaga	tcacgcagcg	19740
cctccggact	gcgcgccgat	agcggaagca	ggtacggtgc	gggcgtactg	gacgcggcct	19800
ggtgcggcgc	ctgctcgatg	agcacgtgcg	cattcgtacc	gctcaagccg	aacgagttga	19860
tgccggcgac	gcgcgccccg	ccgggtgcaa	ccggccaggg	ggtgagccgt	gccgggattt	19920
cgaggggaag	cgtgttccaa	tcgatgtgcg	ggctggcggt	ggtcagattc	agatggggcg	19980
gaatggcttc	gttctgcagc	atcagcgcca	ccttgatcag	tgccggccacg	cccgtgcgcg	20040
cctcgagggtg	gccgaagttg	gtcttcaccg	acccgagctt	cagcttggtg	ccgttggtgc	20100
gccccgctcc	cagcgcgggc	gcaagggtc	cggcttcgat	gggatcgccc	agcggcggtgc	20160
cgggtccggtg	cgccctcgaca	tagctcacat	ccagcgtctg	caagcgcgcg	tctcccacag	20220
cctggcggat	cacggcttcc	tgtgcggggc	cgttcggcgc	cgtcagcca	ttgctgcgtc	20280
cgtcctggtt	gattgccgtg	ccgcgaatca	ccgccatcac	cggatcgcca	tcgcgcagcg	20340
cgtcgagagag	tcgcttcagc	acaaccacac	cgcagccctc	accgcggacg	tagccgtctg	20400
ctgcggcatc	gaatgcctta	cagcgaccgt	cggctgccat	cgccttcagc	ttgcagaagt	20460
agatcgctccg	atccggcgag	agaatcagat	tgacgcgcgc	cgccagcgcg	aggtcgcttt	20520
cacctgagcg	caggctctga	caggcaagg	gcaccgcgac	cagcgatgac	gagcatgccg	20580
tgatcgatcgc	catgttcggg	ccctgcagcc	cgaggatgta	cgagagacgc	ccggcgccaa	20640
cgtcgccgt	attgcccgtg	ccgggtgacg	cgtcgatatg	cgcataccccg	ccgcgcattt	20700
gcaggttgta	ataatcggtg	gaaaagatcc	ccatgaagac	gccgggtccgg	ctccccgcc	20760
gccggtcggg	tggaagccccg	gcgttctcga	tcgcctccca	ggtgacttcc	agaagcagcc	20820
gctgctgtgg	atccaggctg	atgcctcgc	gcggagcgat	gccgaagaac	cgggcgctca	20880
aacgggtcaac	ctgatcgatg	aagccgcgct	accgcgtgta	cattcggccc	gtcgcgccgg	20940
gatccggatc	gtagtaggca	tcgatgtccc	agcggtcggg	tggaacttca	cgtaccgcgc	21000
tgccggccctc	gcgcagcaac	gaccaatagg	catcgagatt	ggatgcgccc	gggaagcggc	21060
agcccgcgcc	gatgagggcg	atgggtcgc	tgccgcgcgt	ctccagctgg	tcgatgcgtt	21120
tctgcacctt	gtcgagcgca	atcacggcgc	ggcgaagctt	gctgagatcg	tctgacctgc	21180
tcattgtttat	tgctctcca	accactggtc	gacctgcgc	agccgcgaat	cgagcagcgc	21240
ttccagttct	tcgcgggcga	ggttctcaaa	ctccggcgct	tccaccgggtg	atgcttcggg	21300
tggaaataacc	gcatggagca	cgtaactgac	gatcgcacgc	agcgacggat	agtcgaacag	21360
cagactcgcg	ggcaaaggct	gccccagtga	ttgggagagc	gagttgcgaa	gttctatggc	21420
cattagcgaa	tcgagtccca	gttcacccaa	aggctgctgt	ggatcgagcg	gtgtggaagt	21480
cgcgatgccg	acaaagcgcg	ccagtgactc	cctgatgtgc	gcaatgagga	tggttccgcg	21540
ctgcgggggt	gtggcttcgt	tcaagcgggt	gcgcagttga	ggtgaaggca	gcgcggcggg	21600
acgcagcaac	tcgccggtaa	tcgagcccg	cggtagcgcg	gcaatctgaa	tggggcattc	21660
atgcaggacg	gcctcgagaa	tgtgtagacc	ctcgtccacg	gagaggctcg	ccacgcgggc	21720
catcgactgg	ctgggtgcgcg	cggccattcc	ggctcccgac	cagcgcccc	agttaatgct	21780
ggtcgccggc	aaacccagtc	cgcgcgggtg	atgcgccagc	gcatcgagaa	cggcggttggc	21840
cgcggcgtag	cctgcctgcc	cggcaggacc	taagagcgag	gatgccgatg	aaaagagcac	21900
gaagaagtgc	agcggcgagat	cgcgggtgtg	atgatggagg	tgtacagcgc	cttcgcgctt	21960
cggcgccatg	acgcttgcca	tccgcgtcca	gtcctgattc	agcagtagcg	cgtcgtccag	22020
cacacccgcg	gcatggataa	cgcgcgcgag	cggtagcgtt	tcgggtgtgga	tgccggcgaat	22080

gagatccgcg	acctcttctt	cccggctgac	gtcgaccgtc	tctgccgtcg	caccaatctg	22140
ttgcagcacg	cgctgctgct	cctcgtttgg	aggccggcgc	ccggccagca	cgacgcgagt	22200
ggcgccgtgc	tccaccatcc	atttcgcgac	tgtaagtccc	agggctccga	gcccgcgggt	22260
gatcaaataa	gtcgcgcccc	aaaccagacg	gactgcccgt	cgcgcgctgg	gtcggcgggt	22320
cagtcgcggc	acgtagcgcc	ggttgcttct	ccacgccgac	tgatcttcgc	cgtcgaaatc	22380
acgcatctgc	gcggccgcgc	cgccgcgcga	agcatgcgca	tcgtcgggat	ccagatcgat	22440
gagcccgccc	cacagatccg	gggtgctcgc	cgcatcacc	cgccgaagc	cccagagcgc	22500
ggcctgcatg	ggattgtgca	ccgcaactgt	cgctgcgcg	ccggccgtta	ccagccatag	22560
ccgcggaccg	gacttcaggg	acttcaccag	ggccagagtg	ctgcggcagc	cgagctcata	22620
atcatcgaga	ctgtacaggt	tgacgatccc	gcgccagtca	cgctcaccga	ctagggacat	22680
gtactcgccg	gctggcggca	cggtaacgca	catctcgccc	tgagctgtga	gcgcatctgc	22740
cagagcgccg	gccgcgccgc	cactgtcggc	caggatcagc	catgccccag	gctgtagcgt	22800
tcgcgaaggc	tggcggagcg	gttcggggcg	ccactcgacc	tcatacaatt	cgggcttccg	22860
ttccgagcgc	tgccccatg	cgcgagtgc	gcgccgaaa	ctcacgccct	gaagttcccc	22920
gagaacgcag	ccctccgagt	ccagcaactg	cgcttcgccg	gtaaagccgt	ccggcgaatg	22980
ccggagaatt	tgcgcgcccc	cccatacggc	gcctccagg	ctgccgtaaa	aacaaacgcg	23040
atcgataccg	agcggagcga	atatcggtat	ttcggcgcca	tccgcaagcg	cgggactcgc	23100
cgcgccgcta	agcaattgca	ggccggcttc	cgccaattca	caacggggat	tgagcggcgt	23160
tgcggaatca	atcgcgccca	gcgcttcctg	ttcaccgaaa	tgaatgcgct	gtatgcggcg	23220
gtagctcggc	cccagttcta	tctcgaggtg	gcgcagcagc	gaatagtacg	tgtctccatc	23280
caccgcaggc	cggcgttcat	cgaccagtgc	gggcacggga	gcgacaccag	cgtgggcggc	23340
aatattgccg	gcagcatgta	agttccacga	gccgtcggac	aagctgagta	tgcggaacga	23400
ggcatgccgg	tcatcgctct	gtgaaagcac	gagctgaaca	gccgtgtcgc	gctccgctga	23460
aaggatcaga	gggtgcgcga	agttcacgtt	ttccagcgtg	tgccggcccg	cgccaaacac	23520
ctccgcccgc	gcctcgagcg	ccatggccag	gaagtacacg	gccggggcca	ccaccgaacc	23580
gtaatatcgg	tggtctgaga	gtagaggcga	agccgtcgat	agtttcgact	cgaagataac	23640
gtctgccacc	ggtagcgaca	gccggcacc	gacgagacca	ctcgcaaccg	ctacaggttc	23700
cggtctggaa	ctccgctcga	tccaatggcg	gcgtctctcg	aaaggatagg	ccggcagggc	23760
gacacgcctt	cgcgaaatacg	gacggtcgaa	ctcctgccaa	tcgatgtcga	acccaccctg	23820
atatagcgtc	gccacactgc	tgagaatcgt	ctccactca	tcgcggcctt	tacgcagcga	23880
cggcagccac	tgcttgggct	cgtcgggcag	gcacttttgg	cccatgcoga	gtagaaccgg	23940
cttaggaccg	atctcgagaa	acacgtcgca	gccttcgtcc	ttgagcgttt	ggataccgtc	24000
ggcgaaacgg	acaggggttt	gagcgtgatc	tcgccagtac	agcggattcg	ccagctgtcc	24060
ctcgccggcc	agtttgcccc	tgaggttcga	aaccaagccg	atcgaaggat	tgcgccacgc	24120
gatcgccgcc	gcccggcggt	gcaggtcgc	cagaatcgga	tccatgctcg	agctgtgaaa	24180
ggcgcgcgca	acggccagca	tctgcgtttt	gatgccctcc	gcacgtagag	ttgccagcgc	24240
gctctcaata	tctgcggcg	caccgaaat	cacgacctca	gcgggtccgt	tgatggccgc	24300
aatggagacg	cgcgaggtga	tcgtgcggc	acagcgtgc	tcgccggcgc	tgaccgcagc	24360
catcgcacct	tccggcaggt	tctgcatgag	ccggccgcgt	tcggcaacta	agccgagcgc	24420
atccggcagg	ctgacggcgc	cggaatata	cgccgcgcgc	tattcgccga	cgctgtgtcc	24480
catcaccagg	tcgggcgtca	caccccagga	cttcacaca	tcgcccaagg	ccactgcaa	24540
agcaaacagc	gcgggctgcg	cgccggcggt	cgcgctgagc	aacgcgtcat	cggccaacag	24600
cgccggcaga	tcgagccgtc	cattcagcag	agctgcgcgt	tcattccatg	cggcgcgaaa	24660
caccggctgc	gactcgtaga	actggcgggc	catgccgcgc	tattgcgcac	cttgcccgggt	24720
gaaaagaaac	gcaatcttgg	ggcgctctg	ggcgatgcga	acccgtcgtg	cctccgtcag	24780
tcgttggcga	gcctcgtcgc	tcgaccgggc	cacaatgcag	atacgggtgc	ggaagtgcac	24840
gcgccttgca	ttggccgtga	atgcgacatc	gccgaacgac	aaaccgggct	ggttgtccat	24900
atggccgcga	tacgagcgca	ccagttcttc	gagggccgcg	tctgtattgg	cggaacaggca	24960
aagcacatgt	gcggatcgtt	cgggcgcgac	tcggccgggc	gtcaccggcg	gcgcttgctc	25020
cagaatcacg	tgagcgttgg	tgccgccgat	cccgaacgaa	ctgactgccg	ctcgtctcgg	25080

```

gggtctttccg gcggggccagt cgagcagccg cgtactcaca cgaaacggag tgtttgcgaa 25140
atcaattcgc ggattcggac gctggaaatt caggctggga ggaatctggc cgcgatggac 25200
ggcaagcacc gtcttgatca gcccggccac accggccgcg acgtctagat gaccgatgtt 25260
gggtcttgacg gatccgatat acacatcgcc gcttccgttt ttcggaaagt tggcagcgat 25320
ggcggcgatc tccaccggat cgccgagcgg cgtggctgtt ccgtgggctt cgatgtagcc 25380
gatggactcc ggcttcacgc ccgccatctc ttgagtgcgc cgaatcaatc gcgtctgacc 25440
gtccacacct ggagcggtaa accccatgcg ctccggcgca tcattattaa tagccgctcc 25500
gcgaatgacg gcgtagatcg tgcgccatc ggccagagcg cggctcaagc gcttgaggac 25560
gaccacaccc gcgcggttg ccggcaccgt gccttgagcg gactcatcga aggcgcggca 25620
gcgcccgtcg ggcgacagga tcatgcccgg ctggtgcagg taccacacgg actgcggaac 25680
attgatggca actccccgg ccaaggcaat gtccgagggc ccgcgtgca agctctcgca 25740
tgccatcacc accgacacca gcgaggtgga gcacgcgcgc tgaaccgtca ggctgggccc 25800
gcggaggttc agcttgtaag agacacgcgt ggccaggaaa tccttgctcg tggccgtcag 25860
cagctggtac gcggaggggc gtgagaaatc gaacggctcc gcggtggcga ggttggtcag 25920
caggtaggta ttgacgcgc atcccgcgaa aacgccgatc gaacccttat agcttcgcgc 25980
cgcatatccc gcgttctcca tcgcttccca cgcgcactcg agaaacacgc gatgctgcgg 26040
gtccatgata tccgcttcgc gcggactgta gccgaagaac gcggcatcga aaaactcgat 26100
gccgtccagc agacccttgg ccggcacgta gctcgggtcc tggaaagacct ccgggctgat 26160
gccgcccgc agcagatctt ccggcgaaag cctggcgatg gaatccacac cgtcgcgcag 26220
attgcgccag aactcctcca cattgcgcgc cccggggaac cggccggcca tcccgataac 26280
tgcgatccga tcttctgcga ccgcagccgc aggttccgca gcggcgggtt cggatttttc 26340
tgccaggccg gcaagcgact cgatcgtcgt atgccggaac agatcgacga cggagagcgt 26400
caaccccagg cgctcctcga gcagtccgcg caccgtgtg agcattagcg agtgcgcgc 26460
gacatcaaag aagttctgcc gatagtcgac gtgctccacg cgcagaactt cacgccagat 26520
ggacgcaatc gtctccacca catcgccgcg catcggtcgc cgagcagcaa ccggcgttgt 26580
gggcaaaccg ggaagcgctg tcgcgtcgat tttgccgttg ggcgtcagcg gaaggaggga 26640
caggctgaca aacgccgagg ggatcatgta atcgggaagg cgcgttgcca gccacgaccg 26700
caaatcgctc tgcagatcgc gcacgtcgcc cgttgccgga acgagatagg cgatcagccg 26760
atcgtccttc acgaccgtaa tcgcctgctt cacggcaatg tgcgtctcga tcgcggcctc 26820
aatctcggcc ggttcgatgc gaaacccgcg cagcttgatc tggcgatcga ctcgtcccag 26880
gcactcgact gcgcgctcgg aacggtagcg agccagatcg ccggtagagt aaatgcgtcc 26940
tcgatcacgc cactcgcgga atttctcacg cgtgagctcg gggttgcgat gatagccccg 27000
cgccagtccc gctcctccga tgtacagctc tccgggaact ccggggggaa ccggctccat 27060
gcgcgaatcc aggatgtata actgcgtgtt gtcgatggga tggccgatcg gcacgatgct 27120
atcggaggca cccagctctt gtgtcttgtg cacggccgac catatggtgg tctccgtcgg 27180
tccgtaaaga ttccacagct ctacgccact atcgagaatg cggcgcgcca gttccggcgg 27240
cagagcttca ccgccgcaga aaacacggaa gcctttaccg ggcttccagc ccgaatccag 27300
caattgcgcg caaccgctcg gggtcgctg catgaccgta gcgcccgaact tatccagcag 27360
ggtggtgagc cgctcgccgt caaccacgat ctgcggggtg gcgacgatga cgcgggcgc 27420
ggtgatcaac ggcagccaga tctccagtcc ggcaatatcg aatgacacgg tggtgacggc 27480
gaccagccca tcggcggctg tcagacccgg ctgcgcgtgc atggagcgca gcagattgac 27540
tagcgacgag tggcggatct ccacgccctt cggtcgcccc gtcgatccgg aggtatatat 27600
gatgtaggcg agatcgtcgg gcttgctgcc gctgacgaga ttcgcagctt ctggttcgac 27660
ggcgaccgcc atcatcgcca tcatcgccat catctcagcc accgcctcct gcgtgaggac 27720
cgctgcggt tgcacttcat cgagaatccg ggcgagacga tccttggggg gcgcgggatc 27780
gagaggcagg tacgcgctgc cggacttcag aatcgcaagc agcgcaatca ccatctccag 27840
cgagcgctcc atcgccagag cgatgatctt tcccggggccc gcgccggatg cgctcagacg 27900
atgagccagg cggttggccc gcgcattcag ctccggcgtag gtcaactgat ggtcttcgaa 27960
gacaacggcg acggcgtgcg gagtgcgttc cgcctgagct tcgaccagtt catgcgcaca 28020
cccgttcgga ccggcatcgc gccgtgtcgc attgtgctgc tcgagcatcc ggcttcggac 28080

```

cgcggggggac	aacagcgcag	cggttgaaat	gcgagcgtcg	ggatccgtca	ccacgctcgc	28140
cagcaggggtt	cggtagcat	cgagcagggg	ggcgatgggt	gccgcatcga	acaaatcggg	28200
gttgatttcg	gcggacgcca	tcagtcctac	gccggatggc	tcgaggggtca	cgccgaggtc	28260
gagtttgat	ccgccgttgt	gcatgtactc	gcgcgagatg	gtgagcccag	gcatgacggg	28320
gatggccggc	gcatcgggca	gcagcgcgaa	ggagacctga	aatacaggcg	accggctcag	28380
gtcccgcgga	ggatgcagtt	cctcaaccag	gcgttcgaaa	ggaaagtcct	gatgagagag	28440
ggcgctcaaa	gcggtgtcgc	gggtgcgggc	gagaagactg	cgaaacgacg	gatcgtcgcg	28500
cagatcgccg	cgaggagcga	tcattgttggc	gaaacaaccg	acgagacctt	ccgtttctcg	28560
ttgtgtacgg	cccgcgactg	gaaccccgat	aaggatgtct	tcctgcgcgg	tatagcgatg	28620
cagcagcacc	tgaacgccc	cgattgcctg	catgaacacc	gtcgtctcct	cacgcaaggc	28680
aaacgcgtgg	agtccatcgg	tcaaatacag	gccgaggggt	gtggtctcca	cgccgccccg	28740
ccaggtctgc	tgcgcggggc	gggggcgact	ggtaggaaag	tcgaggaaaag	gcaagggtgcc	28800
cgacagctgt	ttcttccagt	actgctgcgc	ggtttggttc	agcgacgtct	gctgatggac	28860
ggcccagtcg	ccatactgaa	tcggcagttc	catgagcggc	gatggccggc	cctgcacgaa	28920
cgcttcgtac	gatcgcgtca	ggtcgcggac	gaacgtctcg	accgaccacg	catccgcgat	28980
gatgtggctc	aacgtcagca	ggagaatctg	ctgcttgtca	tcgaggcaga	tcagcttggg	29040
ccgcagaagc	gggggttttc	gcaggtcgaa	cgggatctgg	gcatcacgca	aggccatttg	29100
ccgcgcttct	gcgattccgt	cagcctgaac	aaccggaagt	tccagtgtca	ctcgcgccag	29160
gaggtctctg	cgcgctcttc	catccacacc	gccaatgcag	ctgcgcaggc	tctcgtgccg	29220
ctgcaccacg	gcctccagac	tccgcaggag	gacgcgaata	tccagcggac	ctcggatatg	29280
cagcgctatg	ggaatgttgt	aggcgggaga	atccgggtcg	agctgatgga	gaaaccaaag	29340
ccgtgctctg	gccagcgaca	agggtgcggc	atcccggttt	tcacgcccg	ggatgcgatg	29400
ttcggggctg	ttttcctgca	gcagacggtc	gagcaattgg	cgccggggcg	gcgagaggtc	29460
tatggatatt	ggcgacgaat	tctgcattac	aaccgcgtgt	gttcctagtc	ttggggcgcg	29520
ctcatcatac	gctcgatttg	aacatctgac	atltgggaaa	cagcgatcag	caaatcggcg	29580
gctcgctctg	cccatcctgg	gtctacgccg	tctgaatag	cgacggcgaa	gcctcgaacc	29640
gtgggggctg	taaacacggt	tcgaaagggc	acttccacgt	ggagcatgtc	gcgcacgcgg	29700
gcgatcatct	gcgtgaccag	cagcgaatgt	cctccagagt	cgaagaagtg	atcatggacg	29760
ccgatgccat	ccatgccgag	cacctcgccc	caaatgtggg	cgagtacctg	ttccaccgga	29820
gtttccggag	gcgtgaatgc	ttcggcggtg	gctcgccggc	tgggctcggg	atcgggcagg	29880
gcgttacggg	cgatttttcc	gttgggcgtc	agcggcattt	cgtggagcac	gaccacgcgc	29940
gtcgggatca	tgtagtcggg	cagcttctcc	ttcagatgag	tacgcaactg	cggcacaacc	30000
gtgcgcgtat	agacggctcg	cagcggatcg	ttcgtatacg	ggccggccag	gcggcgctgc	30060
ggacgggaag	ccggcggacc	ggccgcgcga	cggcagaagg	tcgcgtcgaa	gcgtccgtgt	30120
ggcccatgac	tgtctcagtc	gattgccacg	cggtacggca	ggtcttcgtc	catacgccat	30180
agatcggcgg	gatcgacgcc	ggaaggcgac	gtctggcgca	gccgggtccc	caactccccg	30240
agtgtctctg	gagcttcgtc	accgttcatc	caggtcacia	tggcgctttc	ggcgggtcaac	30300
cgtgcgttcg	gaatctcggt	aaatgcggcc	aactccgggt	gagcgtccgt	cagtactctg	30360
cgtatttcgg	ccgcggctct	gcaacgcctg	cgatccgatt	ccggctcctc	cgcttcccgc	30420
gatccgatat	gcaggatcgc	ctggtagcgg	aagcgggtca	gctcgttatg	cgaccggccg	30480
cgacgcggca	ggattttcaat	ccggccgatc	tccggaatct	gttcgcggag	agcaaagaag	30540
aacgcgggat	cgaccacgag	ttcctcttcc	tgcgacgcga	gcgaacgcac	gcgttgccga	30600
aactcattcc	gggtcaacga	cgcgggtgcg	cgctgaactt	ctaaagaagc	gtaaaacgtc	30660
tccagcagcg	ggagactgcg	gacatcgccg	acaaatacga	tgccgcccgg	tttgaccaca	30720
cgcaccgcct	cgccagcac	gcgcgcgaga	tacgttctgc	cggggaagta	ctggataacg	30780
gagttcagaa	caaccgcate	gcacgagcga	ctgtcgatct	cgcacgcgtc	gtcggccgcc	30840
tgccggaacg	tgcggacatt	tgccaggccg	gtgcgggtccg	cgtgagcggc	gatgtagtcc	30900
agcgccttct	gcgaaaagtc	cgtggcccag	tactccgaac	agtggggagc	gacgcggaag	30960
agcagcagtc	ccgtaccaca	gccaatctcg	agcacgcgac	gcggccgcga	ggccaggatg	31020
cgatcgacgg	aatcctgcac	ccactcccgc	atctcggcag	ctggaatcgg	ctctccggta	31080

acactgcttc tccagccgac gatgttgaac tccggatccg cgttcggcgc attctgttca 31140
 tatgtggtgt cccagacgga ttgccactgc gtcacgtgct cggactcgac tccgtcgtgg 31200
 aatgtgtcgg cggctgcgt cgcgcgatgc ccgtcagcaa gggggacaat gtaggccgcc 31260
 agatacttac cggccgcgtc attttctctg gcggtgacca cagcatgtcg gaccgccggg 31320
 tgactgcgga ccgcggcctc gatctcgcgg gtttcgatgc ggaacccgcg tatcttcacc 31380
 tgggtggtcga tccggccgag atactcgagc gcgccgtcgc gttggcggcg ggcgagatct 31440
 cccgtgcgat acagccgagt gccatgaggg tcgaacgaat tggcgacgaa cttgtccgcg 31500
 ctgagttccg gacgattcag gtatccacgg gcgagcccg cgcgcccgat gtacagttcg 31560
 cccgcaacac cgatgggtgc gggctgcac cgatcgtcaa gcacatagag ctgagtggtt 31620
 gcgatggggc ggccaatcga aaccggtcgg tcacctgtcg tcacccggtg gatggcggac 31680
 caaattgtcg tttcggtagg tccgtaaaga ttccatagcg ccgcgggtcg ttgcaggagc 31740
 cggtcggcaa gatcgcgagg aagggttca ccgccgcaga gcgccgtcag gcggcggtcg 31800
 ccgggccagc cggatgcgag cagcagacgc cagggtggcg gagttgcctg catcattgtc 31860
 gctttgtcgc gcgcgagttc cctcgccagc ctctcaccat cgacggcgt ctcttggttc 31920
 gccaccacga cgcgcgcgcc ggcgtcaag ggcaaaaaga tctcgagcgc ggaaatgtcg 31980
 aacatgaacg tcgtgagggc gagcagcgta tcgcggtcgc tgatgcccg ctcatgccgc 32040
 atcgacgaaa gaaaattgac gacggcctgg tgtgtgatt gcacgcctt cggccggccg 32100
 gtcgaaccgg aggtgtacag gacataggcg agatcgccgg gagtcgcgag cgggttcgga 32160
 ttggtgtcgg gctgcgtcca tacttccgat tccgtgacat tcagcacaac gaccggcctg 32220
 gtctcttcca gcatcagccg aagacgttgc gccgatatt ccggttccag cggaacgtag 32280
 gccgcgccgg ctttcaggac gcccaacagc cctgcgacgg tttcgagcga ccgcgtgaca 32340
 tggatgccaa ccatttcgcc ggggtccagc ccgcgcgagc ggagatagtg cgcgatccgg 32400
 ttggcgctcc cgttgagttc gcgatatgtc agattctgct caccgaagct caacgcgatg 32460
 gcgtcggggc tcaactccac ctgagcttcg aacagctcgt gcacgcattg ggacgggaat 32520
 tccgcggcgg tcgcattcca ctcttcgagc agctggatgc gttcccgggt tgtcagcagc 32580
 ggtagatcga caactggaca ggccgggattc tccgcgattc cttccagcag caccggcgaag 32640
 tgcaaggaga gacgttcaat cgtggcagca tcaaaaatgt ccgtgttgta ttgcagaaag 32700
 gcggagaggg cttcatcggt ttcgaccatc atcagatcca ggtcaaaccg gctctgtcgc 32760
 agcggcatcg ccagggactc cagtgtgagg ctgccccagg ccagtcgacc gccggactga 32820
 cccaacatga acggcacgga ttcgggaatg cgatgaggct gctggagcac gaatagaacc 32880
 cgcagtcggg gacccaaccg ctccacgatc cgggcatacg ggtactcctg gtgctcgatc 32940
 gcgccgagaa gcgtttgccg aatccggggc agcacgctat tgaaatccgg atcgccgtgaa 33000
 agttctcttc gcaggattac gggattcacg aagtatccga cgagatcggc gaattccggg 33060
 tgcgtccgac cgttggtgag ggtgccggtc aggatctctt cttgtgaggt ccaacgggag 33120
 agaagcactt gaaacgccgc catcagcgtc gcatgcagcg tcgcgttctg ccgccgcgcg 33180
 agcgccttca gtttcgcagt cagcgcgggt tcgattcgga acgagtgaga gtttccccgg 33240
 aaactctgca ccggcggact gggacgatcc gacgggagat tcagaaccgg aagctggccg 33300
 gaaagctgcg aggaccagta gttccaaagc cgctcgccct cggttccggc caacagttcg 33360
 ttctgccagc ggacgaaagc ggcgaaagtc gcgaccggcg gcgcgacagg cggaccgcca 33420
 gctgtcctcg cgaggtagat actgcggagt tcatccacca tcaccagcag tgaccagaag 33480
 tcggcgagga tgtgatgcac cacgatggcc agaacctgat ccttccccga ctgcaccagg 33540
 agacgcgagc ggaaacagtt ttcgccgaga ttgaaggggc cgtggaagac gccgtcgatc 33600
 agcacgcct catcgtccgg cgaacacggg atcacttcga aatccaccgg gacgctgctg 33660
 tggaccgttt gaacgggtgc gccgccactc tccgcaatcg tcgttcgcag gcccgatga 33720
 cgatccacca ggtcctgcag cgaacggcgc aacgcctgcg gatcgaaagc gcctctcgcg 33780
 cgcgcgatcc acgcgatgtt gtatgcggga ctttccggcg cgcttcggtg aataaaccaa 33840
 agcgcctgct ggccggcgct gagagggtag gagaggcgag gaaccgaggc ctgcgccgca 33900
 ggttccggcg ccaccgtcgt gcgttcgctg aggcgcctta gatcgcttag atccctggcc 33960
 agttccgcaa cgctggggcc gtctagaaat cggaccatgg gcagcaagac gcgcagatcc 34020
 gtatcgatcc gggtgcgtaa ttgcaccgcc atgagcgagt ccaatcccat acgcaccagc 34080

```

ggctgctgta agtccaccgc cgattccggg cactgcagtt tttttttttt tttttttttt 34140
tttttttttt tttttttttt tttttttttt tttttttttt ttttaaatgcg gtagttttatc 34200
acagttaaat tgctaacgca gtcaggcacc gtgtatgaaa tctaacaatg cgctcatcgt 34260
catcctcggc accgtcaccc tggatgctgt aggcataggc ttggttatgc cggtagtgcc 34320
gggcctcttg cgggatgatc cctgtcagtc atgcgggcaa cttagccgag ccctacgaca 34380
ccgcccgtgg gaaggtgagt gtctaactgc gtgacaacgc cagcgcacag cggcggacaa 34440
ccgcgagcac ccatggactg gcgcgcgagg tgagaagcac actggcccaa ggtcgagcgc 34500
ccaccaagt tgcttcggga cgaagaggtc gtgggttcaa atcccgcac cccgacagag 34560
aaacaccagg tgaggcagac cgtaacgtta cggctctgct cacctgtttt ctgtgcgtgt 34620
ctatctgcgt gactatcgcg ccggaccccg cttgaagatg ccgtccatga ccacagcgcc 34680
ggtctggatg acgggcccga tctgcttcgg gtagacctcc tcagtcacgg ccgtaccgga 34740
gtgtccgacg agccgggaga tctcctccag cgggacgccc cggtcggaca gcagggacac 34800
gaagctgtgc ctgagctccc tcgggtgtcca ctgctcggcg ttgatcccg tggcctcctt 34860
gagcgccctg cggaaggcgc gccggacgtt agtcgcgtcg agcggcttgc caacggccga 34920
cgagaagacc aggcctgtgt cctcccactt gtcaccggcg gcgagccgtt cccagccctg 34980
gtcctcaaag tgctgccaga ggacctccac gcaacgcgcc ggcaggggca gcgttcgccg 35040
agacttcagg gttttcgtgt cccacccgcg ccggaccgag cgccagacgg cgatgtgcgg 35100
aggctgcggc ggctcaacgt ccggacttcc cttgaggaag acgtggctcc aggtcagcgc 35160
ccgcagctcc tcgggtgcgcg caccggtcag cagggcgacg acgatgtagg cgtgcacga 35220
cgtgccctcg gcagattca gcaccgctc ggctggggcg aaggtgagcg ccttggacgg 35280
ccggccaggc tggccctggg gcacagagca cagctccacc acgttgcgct tcacctgtc 35340
acgcgccatg gccgccttga ccgcccgggt caggcaggag tggaccgcct gaaggctgcg 35400
cgtgctcaga gtctgagcct tggcggccag ccagcggctc acgtcctctg cgctgaggtc 35460
acgcagcttc cgggcaccca aaccgggtat gacgtgcttc tggcttaggt ggggtgcagt 35520
ctcgacgggt cgctgggtcac ggccagcgag accgtaggca agccagtcgt tcaccgcgtc 35580
ggcgacgggt taccctgtgg gtgcgatcgc gagaccgtct tcgtgggtcac gcagaacctc 35640
tttgagcttg ttcttagcct ccgtcttgggt cttgccactc ccccgcttga cgatccgctt 35700
accgctcgga tcgaagccga ggctcgccgt ggcgatccag cgctgtctct tctcgtccca 35760
gtggaggccg ccgtcacccc ggctacgtcg cttggccatg gatcgatccc ctgcccggca 35820
aaatagagtg ttctcttgcc ctcttttagca ttcagtgtat ccattaccgt catcaattgc 35880
tcactcccgg ggcgcggtgc gttgtcatcg aataaattga gctgcgcgac tccctgactg 35940
aagaaatccc ccagcatcac gcccgctttt tggtaacgat ggcccgcttg ccagatggca 36000
tcagagatc gcgtagcagc gttaatgata tccctgctgt cctgagtggg cgtcagcagt 36060
ttaccgacg cgctattgcc gtaataaggt tcattgagcg caaatggtga cgtcttaata 36120
aacgtggaga taaaccgaca atattgatgc tcgctgcgaa gtttttcgcg cggccgggca 36180
gcgtaactac aaatggcctg ccgcacgcag ggataatccg tgatgcgttc accaaacgag 36240
cgggaacaga taatttcctg cttcgtcggg gcaaaactct ccagttgcaa acagggttcg 36300
ccgcgcagtt cagcacccgt tctttcgagc acgacattaa aatgtttacg gataaacggg 36360
atatctgtat ccgccaaatc gagaacggtt ttgatcccca tcgcgtccag ttttttgctg 36420
atccgcccgt caatccccca gacgtcatcc acggggagag cagacattaa tttacgctgg 36480
cgttccagat ttgataaaat caccacccca cccgtctgcc gctgccattt ttttgccgca 36540
tgattggcaa gctagcttta tgcttgtaaa ccgttttgtg aaaaaatttt taaaataaaa 36600
aaggggacct ctagggctcc caattaatta gtaataaat ctattaaagg tcattcaaaa 36660
ggtcatccac cggatcagct tagtaaagcc ctgcgtagat tttaatgcgg atgttgcgat 36720
tacttcgcca actattgcga taacaagaaa aagccagcct ttcattgat atctcccaat 36780
ttgtgtaggg cttattatgc acgcttaaaa ataataaaag cagacttgac ctgatagttt 36840
ggctgtgagc aattatgtgc ttagtgcatc taacgcttga gttaagccgc gccgcgaagc 36900
ggcgtcggct tgaacgaatt gttagacatt atttgccgac taccttggtg atctcgctt 36960
tcacgtagtg gacaaattct tccaactgat ctgcgcggat cgatccttgc cgagctggga 37020
tggaagcccc gccgacccac cctggaggag atgatcgagg atgccagggc ctttcacgcc 37080

```

cgccgctgct gaggctccgc cgccggggccc gcaccgccgt cggccggccc gctccgggct 37140
 cgcagcagcg ggcttcggcg cgggcccggg gctcccgggc cggccggcg ggctccgccc 37200
 ggcggccgccc gggggccggg ggcggcgccg ggcggcccgg ggcgtcaggc gccggggggcg 37260
 gtgtccggcg gccccagag gaactgcgcc agttcctccg gatcgggtgaa gccggagaga 37320
 tccagcgggg tctcctcgaa caccctgaag tcgtgcagga aggtgaaggc gagcagttcg 37380
 cgggcgaagt nctcgggtccg ctccactgc gcccgcgcga gcagcgccgc caggatctcg 37440
 cggtcgcccc ggaaggcgtt gagatgcagt tgcaccaggc tgtagcggga gtctcccgca 37500
 tagacgtcgg tgaagtcgac gatcccgggt acctcggtcg cggccaggtc cacgaagatg 37560
 ttgggtcccg gcaggtcgcc gtggacgaac cgggggttcg ggcgggccag cagcgtgtcc 37620
 acgtccggca gccagtcctc caggcgggtc agcagccggg gcgagaggta gccccaccgc 37680
 cgggtggctct cgacggtcgc cgcgcggcgt tccgcagca gtccgggaa gacctcgga 37740
 tggggggtga gcacgggtgt cccggtcagc ggcacctgt gcagccggc gagcaccgcg 37800
 ccgagttcgc gggccagggc gagcagcgcg ttccggtcg tegtgcgctc catcgcgga 37860
 cgccagggtg tgcgggtcat ccgggtcatc accaggtagg gccacggcca ggctccgggtg 37920
 ccgggcccga gctcgccgcg gccgaggagg cggggcaccg gcaccggggc gtccgccagg 37980
 accgcgtacg cctccgactc cgacgcgagg ctctccggac cgcaccagtg ctccgccaac 38040
 agcttgatca ccgggcccgg ctccgcgacc agtacggggg tgggtgctct gccgggcacc 38100
 cgcagcaccg gcggcaccgg cagcccagc tctccaggg ctccggcggc cagcggctcc 38160
 cagaattcct ggtcgttccg caggctcgcg taggaatcat ccgaatcaat acggctcgaga 38220
 agtaacaggg attcttgtgt cacagcggac ctctattcac aggttacggg ccggcttaat 38280
 tccgcacggc cggtcgcgac acggcctgtc cgcaccgcgg atcaggcgtt gacgatgacg 38340
 ggctggtcgg ccacgtcggg gacgttctcg gtgggtgctg ggtcgggatc gccaatctct 38400
 acgggcccga cgaggcgac gtgtacgcca cagcttggcg taatcatggt catagctgtt 38460
 tctgtgtga aattgttatc cgctcacaat tccacacaac atacgagccg gaagcataaa 38520
 gtgtaaagcc tgggggtgct aatgagtga ctaactcaca ttacggatca gtgagggttt 38580
 gcaactgcgg gtcaaggatc tggatttcca tcacggcacg atcatcgtg gggagggcaa 38640
 gggctccaag gatcgggctt tgatgttacc cgagagcttg gcaccagcc tgcgcgagca 38700
 ggggaattga tccggtggat gaccttttga atgaccttta atagattata ttactaatta 38760
 attggggacc cttagaggct ccttttttat tttaaaaatt ttttcacaaa acggtttaca 38820
 agcataaagc tatcgtccat tccgacagca tcgccagtca ctatggcgtg ctgctagcgc 38880
 tatatgcgtt gatgcaattt ctatgcgcac ccgttctcgg agcactgtcc gaccgctttg 38940
 gccgcgccc agtctgtctc gcttcgctac ttggagccac tatcgactac gcgatcatgg 39000
 cgaccacacc cgtcctgtgg atctgcctcg ctggcctgcc gcagttcttc aacctcccgg 39060
 cgcagctttt cgttctcaat ttcagcatcc ctttggcgt accattttat gacggcgga 39120
 gagtataaaa gcacctcatt acccttgcca ccgcctcgca gaacgggcat tccctgttcc 39180
 tgccagttct gaatggtagc gatactcgca ccgaaaatgt cagccagctg ctttttgttg 39240
 acttccattg ttcattccac ggacaaaaac agagaaagga aacgacagag gccaaaaagc 39300
 tcgctttcag cacctgtcgt ttcctttctt ttcagagggt attttaata aaaacattaa 39360
 gttatgacga agaagaacgg aaacgcctta aaccggaaaa ttttcataaa tagcgaaaac 39420
 ccgcgaggtc gccgccccgt aacaaggcgg atcgccggaa aggacccgca aatgataata 39480
 attatcaatt gcatactatc gacggcactg ctgccagata acaccaccgg ggaaacattc 39540
 catcatgatg gccgtgcgga cataggaagc cagttcatcc atcgctttct tgtctgctgc 39600
 catttgcttt gtgacatcca gcgcgcaca ttcagcagcg tttttcagcg cgttttcgat 39660
 caacgtttca atgttggtat caacaccagg ttttaacttg aacttatcgg cactgacggg 39720
 taccttgctt tgcgctggct catcacgctg gataccaagg ctgatgttgt agatattggg 39780
 caccggctga ggtgtttcga ttgccgctgc gtggatagca ccatttgca tagcggcgctc 39840
 cttgatgaat gacactccat tgcgaataag ttcgaaggag acggtgtcac gaatgcgctg 39900
 gtccagctcg tcgattgcct tttgtgcagc agaggatatca atctcaacgc caagcgtcat 39960
 cgaagcgcaa tattgtgct caccaaaacg cgtattgacc aggtgttcaa cggcaaat 40020
 ctgcccttct gatgtcagaa aggtaaagtg attttcttct tggatttcag ttgctgtgtg 40080

```

tctggtttca gcaaaaccaa gctcgcgcaa ttcggctgtg ccagatttag aaggcagatc 40140
accagacagc aacgcgccac ggaaaaacag cgcatacaga acatccgtcg ccgcgccgga 40200
caacgtgata attttatgac ccatgattta tttcctttta gacgtgagcc tgtcgcacag 40260
caaagccgcc gaaagttaac ttgtttattg cagcttataa tggttacaaa taaagcaata 40320
gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca 40380
aactcatcaa tgtatcttat catgtctgga tctgacgggt gcgcatgac gtgctcctgt 40440
cgttgaggac ccggctaggc tggcgggggt gccttactgg ttagcagaat gaatcaccga 40500
tacgcgagcg aacgtgaagc gactgctgct gcaaaacgtc tgcgacctga gcaacaacat 40560
gaatggtctt cggtttccgt gtttcgtaaa gtctggaac gcggaagtca gcgctcttcc 40620
gcttcctcgc tcactgactc gctgcgctcg gtcttcggc tgcggcgagc ggtatcagct 40680
cactcaaaag cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg 40740
tgagcaaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcggttttc 40800
cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga 40860
aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct 40920
cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaaagcgtg 40980
gcgctttctc atagctcacg ctgtaggtat ctgagttcgg ttaggtcgt tcgctccaag 41040
ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat 41100
cgtcttgagt ccaaccgggt aagacacgac ttatcgccac tggcagcagc cactggtaac 41160
aggattagca gacgcaggta ttaggcgggt gctacagagt tcttgaagtg gtggcctaac 41220
tacggctaca ctagaaggac agtatttgggt atctgcgctc tgctgaagcc agttaccttc 41280
ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt 41340
tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc 41400
ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg 41460
agattatcaa aaaggatctt cacctagatc cttttaaat aaaaatgaag ttttaaatca 41520
atctaaagta tatatgagta aacttgggtc gacagttacc aatgcttaat cagtgaggca 41580
cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag 41640
ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac 41700
ccacgctcac cggtccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc 41760
agaagtgtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct 41820
agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc tgcaggcatc 41880
gtggtgtcac gctcgtcgtt tggatgggt tcaattcagct ccggttccca acgatcaagg 41940
cgagttacat gatcccccat gttgtgcaaa aaagcggtta gtccttccg tctccgatc 42000
gttgtcagaa gtaagttggc cgcagtgtta tcaactcatg ttatggcagc actgcataat 42060
tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag 42120
tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aacacgggat 42180
aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg 42240
cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca 42300
cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga 42360
aggcaaaaatg ccgcaaaaaa gggaataaag gcgacacgga aatgttgaat actcatactc 42420
ttcctttttc aatattattg aagcatttat cagggttatt gtctcatgag cggatacata 42480
tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg 42540
ccacctgacg tctaagaaac cattattatc atgacattaa cctataaaaa taggcgtatc 42600
acgaggccct ttcgtcttca agaattcgcg gccgcaatta accctcacta aagggatccc 42660
tatagtgagt cgtattatgc ggccgcgaat tctcatgttt gaccgcttat catcgat 42717

```

<210> 114

<211> 34071

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:insert
d'ADN du cosmide a26G1 - brin codant

<400> 114

```

actgcagtgc ccggaatcgg cgggtggactt acagcagccg ctggtgcgta tgggattgga 60
ctcgtcatg gcggtgcaat tacgcaaccg gatcgatacg gatctgcgcg tcttgctgcc 120
catggtccga tttctagacg gcccagcgt tgcggaactg gccagggatc taagcgatct 180
aagcggcctc agcgaacgca cgacgggtggc gccggaacct gcggcgcagg cctcggttcc 240
tgccctctcc taccctctca gcgcgggcca gcaggcgctt tggtttattt accgaagcgc 300
gccggaaagt cccgcataca acatcgcggtg gatcgcgcgcg gcgagaggcg ctttcgatcc 360
gcaggcggtg cgccgttcgc tgcaggacct ggtggatcgt catccggcgcg tgcgaacgac 420
gattgcggag agtggcgggcg caccggttca aacgggtccac agcagcgctc cgggtggattt 480
cgaagtgate cgtgttcgc cggaacgatga ggcgggtgctg atcgacggcg tcttccacgc 540
gcccttcaat ctcggcgaaa actgtttccg ctgcgctctc ctggtgcagt cggggaagga 600
tcaggttctg gccatcgtgg tgcacacat cctcgccgac ttctggtcac tgctggtgat 660
ggtggatgaa ctccgcagta tctacctcgc gaggacagct ggcgggtccgc ctgtcgcgcc 720
gccggtcgcg agcttcgcgc ctttcgtccg ctggcagaac gaactgttgg ccggaaccga 780
gggcgagcgg ctttggaact actggtcctc gcagctttcc ggccagcttc cggttctgaa 840
tctcccgctg gatcgctcca gtccgcgggt gcagagtttc cggggaaaact ctcaactcgtt 900
ccgaatcgaa cccgcgctga ctgcgaaaact gaaggcgctc gcgcggcggc agaacgcgac 960
gctgcatgcg acgctgatgg cggcgtttca agtgcttctc tcccgttgga cctcacaaga 1020
agagatcctg accggcaccc tcaccaacgg tcggacgcaa ccggaattcg ccgatctcgt 1080
cggatacttc gtgaatcccg taatcctgcg aggagaactt tcaggcgatc cggatttcaa 1140
tacggtgctc gcccggttcc ggcaaacgct tctcggcgcg atcgagcacc aggagtacc 1200
gtatgcccg atcgtggagc ggttgggtcc cggactgcgg gttctattcg tgctccagca 1260
gcctcatcgc attcccgaat cgtgcggtt catgttgggt cagtccggcg gtccgatggc 1320
ctggggcagc ctcaactgg agtccctggc gatgccgctg cgacagagcc ggtttgacct 1380
ggatctgatg atggtcgaaa ccgatggagg cctctccgcc tttctgcaat acaacacgga 1440
catttttgat gctgccacga ttgaacgtct ctcttgacac ttcgccgtgc tgctggaagg 1500
aatcgcgagg aatcccgcct gtccagttgt cgatctaccg ctgctgacaa cccgggaacg 1560
catccagctg ctggaagagt ggaatgcgac cgccgcggaa ttcccgctcc aatgcgtgca 1620
cgagctgttc gaagctcagg tggagttgac gccgcagcc atcgcgttga gcttcgggtga 1680
gcagaatctg acatatcgcg aactcaacgg gagcgccaac cggatcgcgc actatctccg 1740
ctcgcgcgcc gctggacccg gcgaaatggt tggcatccat gtcacgggt cgctcgaaac 1800
cgtcgcaggg ctgttggggc tctgaaggc cggcgcgcc tacgttccgc tggaaccgga 1860
atatccggcg caacgtcttc ggctgatgct ggaagagacc aggcgggtcg ttgtgctgaa 1920
tgtcacggaa tcggaagtat ggacgcagcc cgacaccaat ccgaaccgcg tcgcgactcc 1980
cgccgatctc gcctatgtcc tgtacacctc cggttcgacc ggccggccga aaggcgtgca 2040
aatcacacac caggccgtcg tcaattttct ttcgtcgatg cggcatgagc cgggcatcag 2100
cgaccgcgat acgctgctcg ccctcacgac gttcatgttc gacatttccg cgctcgagat 2160
ctttttgccc ttgagcgccg gcgcgcgcgt cgtggtggcg aaccaggaga cggccgtcga 2220
tggtgagagg ctggcgaggg aactcgcgcg cagcaaagcg acaatgatgc aggcaactcc 2280
cgccacctgg cgtctgctgc tcgcatccgg ctggcccggc gaccgcccgc tgacggcgct 2340
ctgcggcggt gaagcccttc ctcgcatctc tgccgaccgg ctctgcaac gaaccgcggc 2400
gctatggaat ctttacggac ctaccgaaac gacaatttgg tccgccatcc aacgggtgac 2460
gacaggtgac ggaccggttt cgattggccg ccccatcgca aacactcagc tctatgtgct 2520
tgacgatcgg atgcagccc caccatcgg tggtgcgggc gaactgtaca tcggcgcgcg 2580
cgggctcgcc cgtggatacc tgaatcgtcc ggaactcagc gcggacaagt tcgtcgccaa 2640

```

ttcggttcgac	cctcatggca	ctcggtctgta	tgcacgsgga	gatctcgccc	gccgccaacg	2700
cgacggcgcg	ctcgagtatc	tcgcccggtat	cgaccaccag	gtgaagatac	gcgsggttccg	2760
catcgaaacc	ggcgagatcg	aggccgsggt	ccgcagtcac	ccggcggtcc	gacatgctgt	2820
ggtcaccgccc	agagaaaatg	acgcggccgg	taagtatctg	gcggccctaca	ttgtccccct	2880
tgctgacggg	catcgcgca	cggcagccgc	cgacacattc	cacgaccgag	tgcagtcgca	2940
gcacgtgacg	cagtggcaat	ccgtctggga	caccacatat	gaacagaatg	cgccgaacgc	3000
ggatccggag	ttcaacatcg	tcggtcgga	aagcagtggt	accggagagc	cgattccagc	3060
tgccgagatg	cgggagtggg	tgcaggattc	cgctcgatcg	atcctggcct	cgcgcccgcg	3120
tgcggtgctc	gagattggct	gtggtacggg	actgctgctc	ttccgctcg	ctccccactg	3180
ttcgaggtac	tgggccacgg	acttttcgca	gaaggcgctg	gactacatcg	ccgctcacgc	3240
ggaccgcacc	ggcctggcaa	atgtccgcac	gttccggcag	gcggccgacg	acgcgtgca	3300
gatcgacagt	cgctcgtgcg	atgcggttgt	tctgaactcc	gttatccagt	acttccccgg	3360
cgaagcgtat	ctgcggcgcg	tgctggccga	ggcggtgctg	gtggtcaaac	cgggcgsgcat	3420
cgtatttgtc	ggcgatgtcc	gcagtctccc	gctgctggag	acgttttacg	cttctttaga	3480
agttcagcgc	gcaccgcgct	cgttgaccgc	gaatgagttt	cggcaacgcg	tgcgttcgct	3540
cgctcgcag	gaagaggaac	tctggtcgca	tcccgcgttc	ttctttgctc	tccgcgaaca	3600
gattccggag	atcgcccgga	ttgaaatcct	gccgcgtcgc	ggccggtcgc	ataacgagct	3660
gaccgccttc	cgctaccagg	cgatcctgca	tatcggtatg	cggaagcgg	aggagccgga	3720
atcggtatcg	aggcgttgcc	agaccgcggc	cgaataacgc	agagtactga	cggacgctca	3780
gccggagtgt	gccgcattta	ccgagattcc	gaacgcacgg	tgaccgccc	aaagcgccat	3840
tgtgacctgg	atgaacggtg	acgaagctcc	agagacactc	ggggagtgtg	gggaccggct	3900
gcgccagacg	tgccttcccg	gcgtcgatcc	cgccgatcta	tggcgtatgg	acgaagacct	3960
gccgtaccgc	gtggcaatcg	actggagcag	tcatgggcca	cacggacgct	tgcacgcgac	4020
cttctgcegt	gcggcgsgccg	gtccgcgggc	ttcccgctccg	cgacgccgcc	tggccggccc	4080
gtatacgaac	gatccgctgc	gagccgtcta	tacgcgcacg	gttgtgccgc	agttgcgtac	4140
tcatctgaag	gagaagctgc	ccgactacat	gatcccgacc	gcgtgggtcg	tgctccacga	4200
aatgccgctg	acgcccacacg	gaaaaatcga	ccgtaacgcc	ctgcccgate	ccgagcccag	4260
ccggcgagcc	cacgccgaag	cattcacgcc	tccggaact	ccggtggaac	aggtactcgc	4320
ccacatttg	ggcgaggtgc	tccgcatgga	tggcatcgcc	gtccatgate	acttcttcga	4380
ctctggagga	cattcgctgc	tggtcacgca	gatgatcgcc	cgctgcgcg	acatgctcca	4440
cgtggaagt	ccctttcgaa	ccgtgtttaa	cgccccacg	gttcgaggct	tgcgcgtcgc	4500
tattcaggac	ggcgtagacc	caggatgggc	aaggcgagcc	gccgatttgc	tgatcgctgt	4560
ttcccaaatg	tcagatgttc	aaatcgagcg	tatgatgagc	gccgcccag	actaggaaca	4620
cagcgsgttg	taatgcagaa	ttcgtcgcca	aataccatag	acctctcgct	cgcccgccgc	4680
caattgctcg	accgtctgct	gcaggaaaac	agccccgaac	atcgcatccc	gcggcggtgaa	4740
aaccgggatg	ccgcaccctt	gtcgctggcc	cagcagcggc	tttggtttct	ccatcagctc	4800
gacccggtat	ctccgcctta	caacattccc	atagcgctgc	atatccgagg	tccgctggat	4860
attcgcgctc	tctgcgag	tctggaggcc	gtggtgcagc	ggcacgagag	cctgcgagc	4920
tgcattggcg	gtgtggatgg	agaggcgcg	cagagcctcc	tggcgcgagt	gacactggaa	4980
cttcgggttg	ttcaggctga	cggaaatcgca	gaagcgcggc	aaatggcctt	gcgtgatgcc	5040
cagatcccgt	tgcacctg	aaaacccccg	cttctgcgga	ccaagctgat	ctgcctcgat	5100
gacaagcagc	agattctcct	gctgacgttg	agccacatca	tgcggatgc	gtggtcggtc	5160
gagacgttcg	tccgcgacct	gacgcgatcg	tacgaagcgt	tctgacagg	gcggccatcg	5220
ccgctcatgg	aactgccgat	tcagtatggc	gactgggccc	tccatcagca	gacgtcgctg	5280
aaccaaacccg	cgcagcagta	ctggaagaaa	cagctgtcgg	gcaccttgcc	tttctcgac	5340
cttccctaccg	atcgcccccg	gcccgcgcag	cagacctggc	ggggcgccgt	ggagaccaca	5400
gccctcgccc	gtgatttgac	cgatggactc	cacgcgtttg	ccttgctga	aggagcgacg	5460
gtgttcatga	cggcaatcg	ggcgtttcag	gtgctgctgc	atcgctatac	cgcgagga	5520
gacatcctta	tcgsggttcc	agtcgsgggc	cgtacacaac	gagaaacgga	aggtctcgtc	5580
ggttggttcg	ccaacatgat	cgtcctgcgc	ggcgatctgc	gcgacgatcc	gtcggttcgc	5640

```

agtcttctcg cccgcacccg cgacaccgct ttgagcgccc tctctcatca ggactttcct 5700
ttcgaacgcc tggttgagga actgcaccc cgcggggacc tgagccgggc gcctgtattt 5760
caggtctcct tcgcgctgct gcccgatgcg cgggccatca ccgtcatgcc tgggctcacc 5820
atctcgcgcg agtacatgca caacggcgga tccaaactcg acctcgcggt gaccctcgag 5880
ccatccggcg atggactgat ggcgccgccc gaatacaaca ccgatttgtt cgatgcgcca 5940
accatcgcc cctgtctcga tgcgtaccga accctgctgg cgagcggtgt gacggatccc 6000
gacgtccgca tttcaaccgc tgcgtgttg tccccgcgg tccgaagccg gatgctcgag 6060
cagcacaatg cgacacggcg cgatgccggg ccgaacgggt gtgcgcata gaactgctgaa 6120
gctcaggcgg aacgcactcc gcacgcgctc gccgttgtct tcgaagacca tcagttgacc 6180
tacgccgagc tgaatgcgcg ggccaaccgc ctggctcatc gtctgagcgc atccggcgcg 6240
ggccccggaa agatcatcgc tctggcgatg gagcgctcgc tggagatggt gattgcgctg 6300
cttgcgattc tgaagtccgg cagcgcgtag ctgcctctcg atccccgcga ccccaaggat 6360
cgtctcgccc ggattctcga tgaagtgcaa ccgcacgcgg tctcacgca ggaggcggtg 6420
gctgagatga tggcgatgat ggcgatgat ggggtcgccg tcgaaccaga agctgcgaat 6480
ctcgtcagcg gcagcaagcc cgacgatctc gcctacatca tatatacctc cggatcgacg 6540
ggcgaccca agggcggtga gatccgccac tgcctcgtag tcaatctgct gcgctccatg 6600
cagcgcgagc cgggtctgac agccgccgat gggctggctc ccgtcaccac cgtgtcattc 6660
gatattgccg gactggagat ctggctgccg ttgatcaccg gcgcccgcgt catcgtcgcc 6720
acccgcgaga tcgtggttga cggcgagcgg ctcaccaccc tgctggataa gtcgggcgct 6780
acggtcatgc aggcgacccc gagcggttgg cggcaattgc tggattcggg ctggaagccg 6840
ggtaaaggct tccgtgtttt ctgcggcggt gaagctctgc cgccggaact ggcgcgccgc 6900
attctcgata gtggcgtaga gctgtggaat ctttacggac cgacggagac caccatatgg 6960
tcggccgtgc acaagacaca aagactgggt gcctccgata gcacgtgccc gatcgcccat 7020
ccatcgaca acacgcagtt atacatcctg gattcgcgca tggagccggt tccccccgga 7080
gttccgggag agctgtacat cggaggagcg ggactggcgc ggggctatca tcgcaacccc 7140
gagctcacgc gtgagaaatt ccgcgagtg gctgatcgag gacgcattta ctctaccggc 7200
gatctggctc gctaccgttc cgacggcgca gtcgagtgcc tgggacgagt cgatcgccag 7260
atcaagctgc gcgggtttcg catcgaaccg gccgagattg aggcccgcat cgagacgcac 7320
attgccgtga agcaggcgat tacggtcgtg aaggacgatc ggctgatcgc ctatctcgtt 7380
ccggcaacgg gcgacgtgcg cgatctgcag agcgatttgc ggtcgtgggt ggcaacgcgc 7440
cttcccgatt acatgatccc ctcggcgttt gtcagcctgt cctcccttcc gctgacgccc 7500
aacggcaaaa tcgacgcgaa cgcgcttccc ggtttgccc caacgcgggt tgctgctcgc 7560
gagccgatgc gcggcgatgt ggtggagacg attgctcca tctggcgtag agttctgcgc 7620
gtggagcagc tcgactatcg gcgaacttc tttgatgtcg gcgggcactc gctaagtctc 7680
acacgggtgc gcggactgct cgaggagcgc ctggggttga cgctctccgt cgtcgatctg 7740
ttccggcata cgacgatcga gtcgcttgcc ggctggcag aaaaatccga acccgccgct 7800
gcggaacctg cggctgcggt cgcagaagat cggatcgag ttatcgggat ggccggccgg 7860
ttcccggggg cgcgcaatgt ggaggagtgc tggcgcaatc tgccgcagcg tgtggattcc 7920
atcgccaggc tttcgccgga agatctgctg gcggcgcgca tcagcccgga ggtcttccag 7980
gacccgagct acgtgccggc caagggctctg ctggacggca tcgagttttt cgatgcccg 8040
ttcttcggct acagtccgcg cgaagcggag atcatggacc cgcagcatcg cgtgtttctc 8100
gagtgcgct gggaagcgat ggagaacgcg ggatatgcgg cgcgaagcta taagggttcg 8160
atcgcggttt tcgcgggatg cggcgtcaat acctacctgc tgaacaacct cgccaccgcg 8220
gagccgttcg atttctcacg cccctccgcg taccagctgc tgacggccaa cgacaaggat 8280
ttcctggcca cgcgtgtctc ttacaagctg aacctccgcg ggcccagcct gacggttcag 8340
acggcggtgc ccacctcgct ggtgtcggtg gtgatggcat gcgagagctt gcagcgcggc 8400
gcctcggaac ttgccttggc cgggggagtt gccatcaatg ttccgcagtc cgtgggttac 8460
ctgcaccagc cgggcatgat cctgtcgccc gacgggcgct gccgcgctt cgatgagtc 8520
gctcaaggca cgggtgccggg caacggcgcg ggtgtggtcg tcctcaagcg cttgagccgc 8580
gctctggccg atggcgacac gatctacgcc gtcattcgcg gagcggtat taataatgat 8640

```

ggccgagc gcatggggtt taccgctcca ggtgtggacg gtcagacgag attgattcgg 8700
 cgcactcaag agatggcggg cgtgaagccg gaggccatcg gctacatcga ggcccacgga 8760
 acagccacgc cgctcggcga tccggtggag atcgccgcca tcgctgcca ctttccgaaa 8820
 aacggaagcg gcgatgtgta tatcggtacc gtcaagacca acatcgggtca tctagacgtc 8880
 gcggccgggtg tggccggggt gatcaagacg gtgcttgccg tccatcgccg ccagattcct 8940
 cccagcctga atttccagcg tccgaatccg cgaattgatt tcgcaaacac tccgtttcgt 9000
 gtgagtacgc ggctgctcga ctggcccgcg ggaaagaccc cgagacgagc ggcagtcagt 9060
 tcgttcggga tcggcggcac caacgctcac gtgattctgg agcaagcgcc gccggtgacg 9120
 ccggccgcag ctgcgccga acgatccgca catgtgcttt gcctgtccgc caatacagac 9180
 gcggccctcg aagaactggg gcgctcgat cgcgccata tggacaacca gcccggtttg 9240
 tcgttcggcg atgtcgcat cagggccaat gcagggcgcg tgcaactccc gcaccgtatc 9300
 tgcatgtgg cccggtcgag cgacgaggct cgccaacgac tgacggaggc acgacgggtt 9360
 cgcatcgccc agacgcgcc caagattgcy tttcttttca ccgggcaagg tgcgcaatac 9420
 gcgggcatgg gccgccagtt ctacgagtcg cagccgggtg ttccgcgccg catggatgaa 9480
 tgcgacgctc tgcgtaattg acggctcgat ctgcggcgcc tgttgccga tgacgcgttg 9540
 ctcgacgcga ccgcccggcg gcagcccgcg ctggttgctt tgcaagtggc cttggcgag 9600
 ttgtggaagt cctgggggtg gacgcccgcg ctggtgatgg gacacagcgt cggcgaatac 9660
 gcggcgccgt gtattgcccg cgccgtcagc ctgcgggatg cgctcggtt agttgccgaa 9720
 cgcgccggcg tcatgcagaa cctgcccga ggtgcgatgg ctgcggtcag cgccggcgag 9780
 cagcgtgtg ccgcagcgat cactcgcgc gtctccattg cggccatcaa cggaccgcgt 9840
 gaggtcgtga tttcgggtgc gccgcaggat attgagagcg cgctggcaac tctacgtgcg 9900
 gagggcatca aaacgcagat gctggccgtt gcgcgcgcct ttcacagctc gagcatggat 9960
 ccgattctgg cggacctgca acgcccggcg gcggcgatcg cgtggcgcaa tcttcgatc 10020
 ggcttggtt cgaacctcac gggcaaatcg gccggcgagg gacagctggc gaatccgctg 10080
 tactggcgag atcacgctcg aaacctgtc cgtttcgccg acggtatcca aacgctcaag 10140
 gacgaaggct gcgacgtgtt tctcgagatc ggtcctaagc cgttctact cggcatggg 10200
 caaaagtgc tggccgacga cgccaagcag tggctgccgt cgctgcgtaa aggcgcgat 10260
 gagtgggaga cgattctcag cagtgtggcg acgctatata aggttgggtt cgacatcgat 10320
 tggcaggagt tcgaccgtcc gtattcgga aggcgtgtcg ccctgcccgc ctatcctttc 10380
 gagagacgcc gccattggat cgagcggagt tccagaccgg aacctgtagc ggttgcgagt 10440
 ggtctcgtcg ggtgcccgt gtcgctaccg gtggcagacg ttatcttcga gtcgaaacta 10500
 tcgacggctt cgctctact ctcagaccac cgatattacg gttcggtggt ggccccggcc 10560
 gtgtacttcc tggccatggc gctcgaggcg tcggcggagg tgtttggcg cgcccgccac 10620
 acgtggaaa acgtgaactt cgcgaccct ctgaccttt cagcggagcg cgacacggct 10680
 gttcagctcg tgctttcaca gagcgatgac cggcatgcct cgttccgcat actcagcttg 10740
 tccgacggct cgtggaactt acatgctgcc ggcaatattg ccgcccacgc tgggtgcgt 10800
 cccgtgcccc gactggtcga tgaacgcgg cctgcccgtg atggagacac gtactattcg 10860
 ctgctgcgcc acctcgagat agaactgggg ccgagctacc gccgcataca gcgcattcat 10920
 ttcggtgaac aggaagcgct ggccgcgatt gattccgcaa cgccgctcaa tccccgttg 10980
 gaattggcgg aagccggcct gcaattgctt agcgcgcgg cgagtcccgc gcttgccgat 11040
 ggccgcaac atccgatatt cgctccgctc ggtatcgatc gcgtttgttt ttacggcagc 11100
 ctggaggcg ccgtatgggg ggccgcgcaa attctccggc attcgccgga cggctttacc 11160
 ggcgaggcg agttgctgga ctcgaggggc tgcgttctcg gggaaacttca gggcgtagt 11220
 ttcggcgcg tcactcgcgc atgggcgag cgctcggaac ggaagcccg attgtatgag 11280
 gtcgagtggc ggcccgaacc gctccgccag ccttcgcgaa cgctacagcc tggggcatgg 11340
 ctgatcctgg ccgacagtgg cggcgccggc cgcgctctgg cagatgcgct cacagctcag 11400
 ggcgagatgt gcgttacgt gccgccagcc ggcgagtaca tgtccctagt cggtgagcgt 11460
 gactggcgcg ggatcgtcaa cctgtacagt ctcgatgatt atgagctcgg ctgccgcagc 11520
 actctggccc tgggtgaagtc cctgaagtc ggtccgcggc tatggctggg aacggccggc 11580
 gcgcaggcga ccagtgcggt gcacaatccc atgcaggccg cgctctgggg cttcggccgg 11640

gtgatcgcgc gcgagcaccg ggatctgttg ggcgggctca tcgatctgga tcccgacgat 11700
 gcgcattgctt cggcgcccg cgcggccgcg cagatgcgtg atttcgacgg cgaagatcag 11760
 tcggcggtgga gaagcaaccg gcgctacgtg ccgcgactga cccgccgacc cagcgcgcga 11820
 gcggcagtcg gtctggtttc gggcgcgact tatttgatca ccggcgggct cggagccctg 11880
 ggacttacag tcgcgaaatg gatgggtggag cacggcgcca ctgcgctcgt gctggccggg 11940
 cgccggcctc caaacgagga gcagcagcgc gtgctgcaac agattggtgc gacggcagag 12000
 acggtcgacg tcagccggga agaagaggte gcggatctca ttgcgcgat ccacaccgaa 12060
 acgtcaccgc tgcgcggcgt tatccatgcc gcgggtgtgc tggacgacgg cgtactgctg 12120
 aatcaggact ggacgcggat cgcaagcgtc atggcgccga aggcggaagg cgctgtacac 12180
 ctccatcatc acaccgcga tctgccgctc gacttcttcg tgccttttc atcggcatcc 12240
 tcgctcttag gtccctgccg gcaggcaggc tacgcccgcg ccaacgccgt tctcgatgcg 12300
 ctggcgcatc accggcgcg actgggtttg ccggcgacca gcattaactg gggcgctgg 12360
 tcgggagccg gaatggccgc gcgcaccagc cagtcgatgg ccggcggtgg gagcctctcc 12420
 gtggacgagg gtctacacat tctcgaggcc gtccctgcgt aatgccccat tcagattgcc 12480
 gcgctaccgg cgggctcgat taccggcgag ttgctgcgtc ccgcgcgct gccttcacct 12540
 caactgcgca ccgcttgaa cgaagccaca cccggcgagc gcgaagccat cctcattgcg 12600
 cacatcaggg agtcactggc gcgctttgtc ggcacgcga cttccacacc gctcgatcca 12660
 cagcagcctt tgggtgaact gggactcgat tcgctaattg ccatagaact tcgcaactcg 12720
 ctctcccaat cactggggca gcctttgcc gcgagctctg tgttcgacta tccgtcgctc 12780
 gatgcgatcg tcagttacgt gctccatgcg gtatttccac ccgaagcacc accggtggaa 12840
 gcgcggaggt ttgagaacct cgcccgcgaa gaactggaag cgctgctcga ttcgcggctg 12900
 gcgcaggtcg accagtggtt ggagacgcaa taaacatgag cgggtcagac gatctcagca 12960
 agcttcgccg cgccgtgatt gcgctcgaca aggtgcagaa acgcacgcac cagctggaga 13020
 gcgcgcgcag cgagcccatc gccctcatcg gcgcgggctg ccgcttcccc ggcgcaccca 13080
 atctcgatgc ctattggtcg ttgctgcgcg agggccgcag ccgggtacgt gaagtccac 13140
 ccgaccgctg ggacatcgat gctactacg atccggatcc cggcgcgacg ggccgaatgt 13200
 acacgcggta cggcggttc atcgatcagg ttgaccgtt tgacgcccg ttcttcggca 13260
 tcgctccgcg cgaggcgatc agcctggatc cacagcagcg gctgcttctg gaagtcacct 13320
 gggaggcgat cgagaacgcc gggcttccac ccgaccggct ggcggggagc cggaccggcg 13380
 tcttcattgg gatcttttc aacgattatt acaacctgca aatgcgcggc ggggatgcgc 13440
 atatcgacgc gtacaccggc acgggcaata cggccagcgt tgccgcggg cgtctctcgt 13500
 acatcctcgg gctgcagggc ccgaacatgg cgatcgacac ggcattgctg tcacgctgg 13560
 tcgcgggtgca ccttgccgtg cagagcctgc gctcaggtga aagcgacct gcgctggcg 13620
 gcggcgctcaa tctgattctc tcgccggatc ggacgatcta cttctgcaag ctgaaggcga 13680
 tggcagccga cggctcgctg aaggcattcg atgccgcagc agacggctac gtccgcggtg 13740
 agggctgcgg tgtggttgt ctgaagcgac tctccgacgc gctgcgcgat cgcatccg 13800
 tgatggcggg gattcgcggc acggcaatca accaggacgg acgcagcaat ggactgacg 13860
 cgccgaacgg gcccgcacag gaagccgtga tccgccaggc tgtgggagac gcgcgcttg 13920
 agacgctgga tgtgagctat gtcgaggcgc acggaaccgg cacgccgctg ggcgatccca 13980
 tcgaagccgg agcccttgcg gccgcgctgg gagcggggcg caccaacggc aacaagctga 14040
 agctcgggtc ggtgaagacc aacttcggcc acctcgaggc ggcagcgggc gtggccgcac 14100
 tgatcaaggt ggcgctgat ctgcagaacg aagccattcc gcccattctg aatctgacca 14160
 cgccagccc gcacatcgat tggaaacgc tccccctcga aatcccggca cggctcacc 14220
 cctggccggg tgcaccggc gggcggcgcg tcgccggcat caactcgctt ggcttgagcg 14280
 gtacgaatgc gcacgtgctc atcgagcagg cgccgcaaca ggccgcgtcc agtacgcccg 14340
 caccgtacct gcttccgcta tcggcgcgca gtccggaggc gctgcgtgat ctggcgcgcg 14400
 cataccgcga cgtggtgaac gacaacccc cgcacacctg ctacacggcg tgcgctcgcc 14460
 gcacttcata cgaacaccgc gcggcattca ccgggacgaa cgcgcaggac ttgatggccg 14520
 ggctggacag ttttctggcg ggcaaccgga accgcgatac cgccacagg tttgtgccgc 14580
 gcggccagaa gcgaaaagtc gttttcgttt tgccgggaca aggatcgagc tggcccgga 14640

tggggccgcga	cctgatggct	tctgaaccgg	tgttccgtgc	cgccatcgaa	gagtgcggcc	14700
gcgccatgca	gccttacgtc	gactggctgc	tgacgcaaga	gttgcagggg	ccgctcgacc	14760
gcacgcagct	gattcaaccg	gccctgttcg	cagtcggggg	cgccctggcc	ggactgtggc	14820
gccattgggg	aatcgagccg	gacgccgtga	tcggccacag	catgggcgaa	gtcgcggcag	14880
cgcacattgc	aggtgcgctg	actctcgatg	aagccgctcg	ggatgatttc	ctgcgcagcc	14940
ggatgctcgc	cggagtacgc	ggccagggag	aaatggctgt	cgtggaatta	gcgctggacg	15000
aggccatcgc	tgccatcgcc	gggcgctcgg	atcgggtctc	gattgccgcc	agcaacagcc	15060
cgcgcagcac	cgtcctgtcg	ggcgacagcg	cagctctggg	cgaactgctg	cgggaactgg	15120
aggcgaaaga	cgtcttctgc	cgtcgcgtga	aagtggacat	tgccctcgac	agccatctga	15180
tggactccgt	gtgcgcggcg	ttgccggggc	tggtgggagc	gcttcagccg	cggccggccg	15240
cccttggcat	gtactccacc	gtcaccggcg	cagcgattag	cggatgaagag	ctgggtttctg	15300
cgtactgggc	tcgtaatctt	cgccaaccgg	tgatgctgtc	gacggccgct	gccgcagccg	15360
cggcgggtgg	tcgatgatgtg	tttctggaac	tgagtcccca	cccgttgttg	gtccagccga	15420
tccaggaaac	gctcgagatg	cgggcagcga	ttgccgctgc	ctcgttgccg	cgcgatgaag	15480
acggaaacct	cgcactgcgc	cggacgctgg	gagcgcgtgt	gactaacgga	gtcactccgg	15540
actggtctcg	tatttatccc	aacggcggcc	aaactcgccg	gctgcccac	tatccctggc	15600
agcgtgagcg	ttattggatc	gatatccgtc	cggcgaggtg	cgagtctcag	gctttgcctg	15660
gccggcgcat	cccgtcgccg	ctgccggaga	tgcagttcga	gtccactgtg	gaggcgaaag	15720
atttcgcgga	tcaccggctg	cacgatgtga	tcgtgactcc	gggagcgtgg	cacctggcaa	15780
tggcgctcgc	cgtgcgcgcg	caaggtctcg	gcgccggggc	tcaccatgtc	gaacacgtgt	15840
cattgacggg	cgcgctgacg	ctgccggaaa	acgatgctgc	caggcaggtt	caactggtac	15900
tccgtcatga	agagggcggc	ggagcttcc	tccgcatcta	cagccgcgag	gattcctgga	15960
agctgcacag	cgaaggcatg	ctgcaggcgg	gcgattccac	ggcatccatc	gatctggatg	16020
cgattcgcg	ccgctgcacg	gcggagctca	cagccgatgc	cttctattcg	cgactgtggg	16080
atcgcgcta	tcacttcggt	cccacettcc	gaaccatcgg	ccccatctgg	cgcggcaacg	16140
gtgaggtgct	ttgtcgcgtg	gacattccgc	tgacggaaat	gcagacgata	gactgctgtc	16200
tgagttgcc	cgcggccctc	gtccatcacg	acgatttgaa	agatgtgcat	gtgccggtag	16260
gtctggaccg	attctcgctc	gctgaagtgc	ccactggccc	ggtctgggga	tacgcggctc	16320
tgcgcccgga	ttccacgggtg	gatgtccgtc	tcgtcacccg	caccggcagc	gtggtggcgg	16380
aattggtggg	gctgcagtcg	agagtcgccc	atagcggcca	gctcggcgaa	tcggagattc	16440
ccacctggac	ggtgcaatgg	accgcgtcgg	ttcgcccgcg	cgatgccaat	gccggcaatg	16500
ctggcggaac	ttggctcgct	atcggcgagc	cggcgattgc	cgagactctg	caaaagcgcg	16560
gccaaacctg	ccgcacggcc	gatacgtgct	cgggtccgcc	gtgccgtcaa	attgtgtact	16620
gtccctcgcc	gcgcacgcac	gacctgcttt	ccgtattgcg	cagcatcgtg	caagcgggct	16680
ggcctgagcc	gcgcgcctg	tggtgctga	cgcgcggatc	tgccgcggtt	ctcaactccg	16740
acaaagatat	tgatattcga	caagcctggc	tgacacggaat	tgggcggacg	attgcctatg	16800
agcatccoga	gctgcgctgc	acgctcgctg	atctcgatgc	gcacagcaac	gactgcgggc	16860
atctcgcgac	gctgatgctg	tcgaatatcg	cagaggatca	agttgcgata	cggcaaggca	16920
cggtatgggc	gcgcgcctc	agtcttcaca	agatcccatc	cgcacccgat	gtggcggttc	16980
gtgccgacgc	aacctatctg	atcacggggc	ggctcggcgg	actcggactg	caggtggcgg	17040
gatggctcgc	cgcgcgcgga	gcgcgccatc	tcgttctgct	gggacgcagc	gagcgtccctc	17100
ggccacaact	ggaagggtgtc	aacgtcaaga	tcattccatgc	ggacgtggcg	gaccggcagc	17160
agctatcgga	tgcgctcgcg	atcatcgatc	gcgacatgcc	gccgttgccg	ggcgtgttcc	17220
atctggcagg	cacgctggcc	gacggcatgc	tgctcaatct	cacgaccgaa	cgcttcgaag	17280
ccgccatggc	tcggaaagta	gccggcgcg	ggaacctgca	cgaactcacc	gccggccggc	17340
cgtggatca	ttttgttctc	ttctcttccg	ccagcgcgac	agtgggatct	cccggccagg	17400
gcaactacgc	cgcgcgcaat	tcatttctcg	acgcgctggc	tcattctgcg	cgcgccagg	17460
gtcttcccgc	cgtcagcatc	gcgtggggac	cgtggacaca	ggttggtttg	gccgcacagg	17520
cgaaccgcgg	agaccgtctg	gcgcgcgcg	gcattctcggt	tattcaaccg	caacagggat	17580
tgcgcgcgct	ctacaaagca	ttgacgcaga	ttcgcccgca	cgtcgcgtgc	atgaacttcc	17640

atatacgcgca	gtggctccgt	tactatccgt	cggccgcatac	gatgtccctg	ctggccggca	17700
tcgcacccgc	ggccgcggac	accaaaccgg	cggccgacat	gcgcagcgag	ctcctggcag	17760
ttccagccgg	gcggcagcgc	cgcgcgcggc	tggaaacgct	gctgatgcac	gaagccggac	17820
acgtgctgcg	cttcgatcca	gcgaaactcg	acggcagagc	gacgtgggt	gatctcggat	17880
tcgattcggt	gatggccctc	gagtttcgca	accgtctgga	agccgggctg	cgcgtcaagc	17940
tttctgccac	cctgatctgg	cgttaccgga	cattctccgc	cctggcgag	catctcgccg	18000
acaagctcgg	cctgccgctg	gaaagcatgg	cgggcaatgc	tgaaccttcg	accgttgctg	18060
ccgttgctac	ccttgctacc	gttggcaccg	cgcggggcga	ggaccggagt	cccgcgctg	18120
cagacgatct	cgacgccgtc	gcaaaccaga	tcgccgggtt	gggggacaaa	gaaatcgaag	18180
ctttgttgaa	acagaagttc	gctcattttt	caggagcctc	cgagtgagtt	cgatatccga	18240
gcgattcccc	aaccttacgc	cgttgcagca	ggcgtacctg	acgtggagc	acatgcagcg	18300
acgtctcgat	gcggccgaac	gcgacgcgcg	cgaaccctac	gcgatcgtgg	gtctgggctg	18360
ccggtttccg	ggcggcgatg	ggccccgatg	gttctggcag	atgttgcgca	gtggagtcca	18420
tgctattcgt	gaggtaccgc	ctggacgatg	ggacgaggag	tcggtcgggc	gcatectgaa	18480
atcgttgaac	cccgccacgc	cggtgaagat	tcaagccgga	tttctcgatt	ccatcgatgg	18540
tttcgacaac	gatttttttcg	gcatttcgcc	acgcgaggcc	gtcagcattg	atccgcagca	18600
gcggctgctg	ttggaagtgg	cgtgggaggc	actggaggat	gcggggcaga	cgatggaagg	18660
gctctccggc	agccgcacgg	gcgtcttcgt	cgggatccac	agccaaagca	gcgactattt	18720
ctggatgcag	accgccgatg	gcgcgcgcac	cgatccgtat	accgccaccg	gcacggcgca	18780
tagcgtgac	gccggccgac	tttctatatt	gctgaacttg	caaggaccca	gcatecgct	18840
cgacacggcc	tgctcgtctt	cgctggcggc	ggttcacatg	gcgtgccaga	gcctgcgcag	18900
cggcgagtgt	acgctggccg	tggccggcgg	agtgaatctg	cgcttctcgc	cggagtttat	18960
gtacgccacc	tcgaagatgg	gaaccgcctc	gcccagcggg	cgctgcccg	ccttcgacgc	19020
ggcggcggac	ggcatcgtgt	tcggagaagg	ctgcggcggt	gtggtgctga	agcgctgtc	19080
cgatgcactc	gcggccggag	accgggtgtg	ggcgtgggtg	cgcggctccg	cggatcaatca	19140
ggatggccgc	tcggccgggc	tcaccgctcc	caatgtcgtg	tctcagcagg	tcgtcatccg	19200
gtcggcattg	gccaatgcgg	gcgtcgcggc	gcagcagatc	ggttacatcg	aagcccattg	19260
cacggggact	ccgctcggcg	atcccacga	gatcgaggcg	ctggcggaaa	ccgtcggcct	19320
cccgcgacct	gtcggcgatg	tgtgcgcggg	cgggtccctg	aaatcgaaca	tcggccacct	19380
ggagggagcg	gcaggcatag	cgggattgat	taaagcgggt	ctcgcattga	gtcacgagac	19440
gataccgccg	agcttacacg	tgagacagct	gaaccggaat	atccggttgg	agggaaacgtc	19500
gctcgacatt	gtgaagggaag	tcgggccgtg	gcccgcgggt	tcgagacgaa	ggtttgccgg	19560
cgtcagcgcg	tttggttggt	ccggcacgaa	cgcgcagtgc	gttcttgaag	aagcggcgcc	19620
gactggtaga	ggcgaagctg	cgagcgggtt	ccattcccga	ccccccgccg	ccgctgcgcg	19680
ggcggctgtc	cccctcgcg	agggggacac	tgggggcact	cccgacattg	caggcactcc	19740
cgacactgca	gacactcccg	acactgcaga	cactcccagc	attgcaggga	ctgcaggcac	19800
tcgggcaact	acgggcattg	cagacgcgat	gtatgtgctt	ccgctgtccg	cgcattggtgc	19860
ggacgaactg	cgtcgggtgg	cgcgggcata	cggggaattg	ctgacagcgt	cgcacgcacc	19920
gagcctgcgt	gatctttgct	acacggccgc	agtccgccgc	acgcatacc	gatgccggct	19980
cgtgtttcc	ggcagaacgg	ctgaagaact	ggcggcgag	ctccagggga	tcacgatccc	20040
ttcccagcga	cggaaagcgg	tattcgtctt	ctcgggacag	ggatcgcaat	ggatcggaat	20100
ggggcgagc	tggatggacc	gcgaaccctg	tattcgcgag	gcgttggaac	gctgcgaggc	20160
cgccatgcgg	ccttatgtgg	actggtcgct	gaaagaagaa	ctggcgaaagc	tcgaccgcgt	20220
cgaggtcatt	cagcctgcgc	tcttcgcgct	gcaggtcgcc	atcgccgcat	tgtggcgctc	20280
ctggggaatc	gagccggatg	ccgtcatcgg	gcacagcatg	ggagaggctg	ccgccgctca	20340
tgtcgcgggt	gcgctgacgc	tcgaggatgc	ggcgcggatc	atctgcagcc	gcagccggct	20400
gttgagccgg	atcagcggcc	tgggcgggat	ggcgatgggtg	gagctgccgc	tcgcggaatg	20460
tgaggccgtg	ctgtcgactt	acacggaacg	actatcgccc	gcggtgtcga	acggacccaa	20520
ctccaccgtc	atctccgggtg	aagtcgaagc	cctggccgag	gtcgtcgcca	cgtggagcg	20580
gcgaggcgtg	tcttgccggc	cggtgaaagt	ggacttcgcc	gcgcatagcc	cgcaagtggga	20640

cccattgtgc gacgaactcc tgcagtcgct cgacgggatt caaccgcggc ccgcgaccat 20700
 acctttttac tccacgggtga cgggcgcgac gctggagacc accagcctcg acagcacgta 20760
 ctgggctcgc aatctgcgat cgccggttct gttctggcag ggcacccgcc atcttgccga 20820
 cagcgggcac gatgtcttcc tcgagatcag ccctcatccc atcctgctgc ccgccatcgg 20880
 cggcaatgcy gcgctgggtc cgtctctgcy ccgcgaccag gacgaacgcg gttccatgct 20940
 cagctcgctg ggcgcctctc atgaggctgg gcacactgtc gcatggcgga ccgtgtaccc 21000
 ttccggcaat tgcgtgcgcc tgcgccggta tccctggcag cgtcgtcgtt tctggctcga 21060
 cgcttcccc gcgcgacacg cgatcacgtt gggcaatccg ctggtgggaa aacgcgtcga 21120
 agcctcgacg caaccgggca ctttcttctg ggagacggaa ctcagtctcg cttccgtgcc 21180
 ttggctggca gaccatcgcy tgcaggcgga agtcgtcttg ccggtactg cgtatctcga 21240
 tatggctctg gccggaactt ccgagacctt cggtgaaagt ccgtgcgtgc tggagcatgt 21300
 gactttcaca cagatgctca ttgtgcgcg cyacggcgagc atgacgttgc agctggccat 21360
 cgcggtcgat agaccgggga tggcgtcgtt tcggatttcc agccggcagg catcgacatg 21420
 ggtcctgcat gcttccgggg acattcgtca gacgcctgcy gatgcacga ccgtcccgc 21480
 ggattctgcy gagacgggtgc agggccgctg cccacacagt gtgcggcgcy cggagctgtg 21540
 gcgtcagatg gcggagcacg gcgtcgagta tggtcgggt ttcgcgcgc tcgagcagat 21600
 ctggagttgt ccagggtgag cgatcgggcy tctgcgtagc tcggaaacgc gttccactgc 21660
 gccggcgttc ctcgatgcat gtctgcagat catcgccgcg gcgtttggtc ccgccgggtg 21720
 aacctggctg cccgcgggca tcgaccggat gcgctggctg catccgcac gttccgtggt 21780
 gtggacgcat gcgcggctgg aaggacctat ccgcgatctg tcgctgctgg acggagaggg 21840
 acaactggtc gccgcacatg agggctctgcy gctgcagcgc ctggatgcgt cggagcgcac 21900
 cgacatgcgc ggctggttgc acgaactgcy ctgggtcgtc cagccgcacg ccgctgcaga 21960
 gccgcgggcy gcgcgagcgg ccgggtcatg gctcattgtc ggcgctgtgg atagecgct 22020
 caccgcacatg ctgcgcgcta ccggcaaccg cgtgacgcag acctcgccgg aaaagctcga 22080
 tgaactccag ccgcgcctcg aggaaatcgt gtttttgctc gagcacgaac cctcatgcca 22140
 ccgcattctg catctcctcc agaccctggg gcgcacgccc tggcgtcaag caccgcgcct 22200
 atggctggtc acgcgcggcy cgacgcgggt cgatggacag atcctgcaag ccggtatcgc 22260
 tcaggcgctt ttctgggggt tgggcccggac cgtgcattac gaacatccgg aactgaactg 22320
 cagctgacg gatctcgatc ccgcggcgcy cgaagaggaa ctctgcacg aactgctgac 22380
 gaacaacggc gagaatcaaa tcgcctttcg cggcgcgcy cgttacgtcg ccgcgctggc 22440
 tcggcacgaa gcggatatgc aaccggccat gttcaaggcc ggcgatcggc cgttccggct 22500
 cgagatcgat gccccggag tctcgaccg gctgcgttg cgggccacat ccgcgcgc 22560
 cccgcaagcc ggtgaagtgg agattgaagt ctgcgcgcg ggctgaact tctcgacgt 22620
 tctgctgcgc ctgcgggtta tgcgcgacga tgcgcgcgg gcgattgccg gcagcccgcy 22680
 cctgggcygc gaatgctcgg gccgtatcgt ggccatgggg aaaggcgtca ccgactttcg 22740
 catcgagatg gaagtgcgtg cccttgcgcc ttgcagttt ggctcgttgc tcaccacgc 22800
 cgcttccgc gttgccttga agccggccaa cattcccgcc gaacaggccg ccgccctgcc 22860
 tatecgctt ctaccgcgc attacgcgt ctgcgcgag gcgcggctgg ccgccggcga 22920
 acgagtcctg attcacgctg ccaccggcg tgtgggattg gcggcaatcc agatcgaca 22980
 gcgtgcgggc gcggagatct tcgctactgc cgggagtcg gaaaaacgag cgtatctgcg 23040
 ctgctgggc atcgcgcatg ttccggattc gcgctcgat gctttcgtgg acgacatccg 23100
 caattggacg aatcaagaag gagtagacgt cgtcctgaat tcgctttccg gcgacatgct 23160
 ggaggcgagc ttcgatctgc tgcgcgatca tggacggttc atcgagatcg gcaagcgcca 23220
 ttactatgcc ggcgcgaagc tggggcttcg ccggttctg aagaacctt cgtacacgct 23280
 ggtcgatttg ctgcgcacgt ccctgaagcy ccgggcatg acccgggagc tgctgcagga 23340
 gatggtcgca aaattcgaa cggaaacctg gcggccctg gaaacgcgag tgacgaccat 23400
 caccgaatcg gtggaggcgt ttcgcaccat ggcgcaggcy cggcacatcg gcaaaatcgt 23460
 catggcgatg cgagattgcg ccaatgcgcc catcgacccc ctacgctcgg cgttcgatag 23520
 cgagggaacc tacttgatta ccggcggaact tggcgggctc ggtcttaccg tcgcacgctg 23580
 gatgatcgga gcgcgcgccc ggcggctggt gctgctgagc cgcgcgcgc cttcaccgca 23640

ggtccagcaa	gccatcgccg	tcatggacgc	agatgtccgg	acggtgcagg	ccgatgtttc	23700
tcagcgcgat	gaactcgagc	gcgtgatctc	ttccatcgat	cgattgcgcg	gcgtgattca	23760
tgccgcagcc	gttctcgacg	atgcgctgct	actgaaccag	acggaagcgc	atttccgcaa	23820
cgtgatggcc	gcgaaaatcg	acggtgcctg	gaacctgcac	ttgctcacc	gcgactgccc	23880
gctcgatcat	ttcgtgctct	tctctccgc	tgcaggactg	ctgggcgcgc	ccgcccaggg	23940
aaactacgcg	gccgcgaacg	cctttcttga	cgcgctggcc	tactaccgga	aggcccaagg	24000
cctgcgcggc	ctgagcatcg	gttgggggtg	gtggctcggag	gtcgggctgg	ctgccgcgca	24060
ggacaatcgc	ggatcgcgcc	tggctttgcg	cggcatggaa	aacctgacgc	cgcaacacgg	24120
cctcgctatt	ctggaacagc	tgctgaacag	ctcggcttgc	cacgtcgccg	cgatgcccat	24180
caatgtccgc	cagtggcgcc	agttctatcc	caaggcgccg	cagtctgcac	tgttcgagct	24240
tttgcatgac	gacgcggcga	gcgaagccga	tgcgccaaac	gcgttgcgcg	cgcggctgca	24300
atcgcccgag	cctcagaccc	gcaggacatt	gctcgaagaa	catctacagc	agcagctggc	24360
gcgcgtgctg	cgcacgcact	ctcaaactat	cgatccccctg	cgcgcgctga	aggaaactcg	24420
cttcgattcc	ctcatggccc	tggagtttcg	caaccgtctc	gaactcacac	tgggtctcac	24480
gtcccccgcg	accctgattt	ggggtcatcc	cacgctggcc	ggtcttgccc	cgcacctggc	24540
gtcgcaaatg	ggactgccgc	tggctgaagc	gcaggccgcg	gctgctgcgg	aaggagacag	24600
ccgcgccatg	aaaactgcac	tcagcggggt	ggacgacatg	tcggaagaag	cagccgtggc	24660
tgcgctccga	ggagcaaggt	cgtgagggaa	aaaattgcgc	ccatgtcgtc	ggtcaaaactc	24720
gcgtatttgg	cgcggaacat	gcggcaaaac	atcgcaggct	tcgacctggt	tcacgccgaa	24780
cccatcgcca	tcgtcggcat	ggcgtgtcgt	tttcggggcg	gcgcgaagaa	tcgggacgcc	24840
ttctggacgc	tgttgaagaa	cgggtgtcgac	gggtgtcaccg	aggtgccgcc	agaccgctgg	24900
aaactcggacc	agtactactc	ctccgatccc	gatgctccgg	gcaaggcgta	tgcgcgatat	24960
gccgccttcc	tcgaacgcac	tgacggtttc	gatgcggaat	tcttcggcat	ctccccccgc	25020
gaagctctga	acatggatcc	gcagcagcgg	ctgctgctgg	aagtgtgctg	ggaagcggca	25080
gaggacgccg	gcactctctc	cggccctctg	gcgggcagcg	cgaccggcgt	ctttgcgcgc	25140
tctgcgccc	aggacttcgg	actgtttcag	tacgccgacc	ctgcccgcac	cggagcttgg	25200
tcgggttcgg	gcgtggcgca	tagcatgttg	gccaatcgca	tctctatctc	gctcgacctg	25260
cgcggtccga	gcattggcgg	cgatacggcc	tgctctcccg	cgctcgtcgc	cgtccatctg	25320
gcttgccaaa	gcctgcgcgc	gcgcgaatgc	gatgcggcat	tcgccggcgg	agtgaacttg	25380
atctgactc	ccgagggcat	gatcgctttg	tcgaaggctc	gcattgttgg	gcccgcagga	25440
cgtgcaaga	cgttcgacgc	cgcagccgac	ggttatgtgc	gcgggcaggg	ctgcccgcac	25500
gtgctgctga	agcggctctc	cgatgcgctg	gccgatggcg	atgccatccg	tgcagtcac	25560
cgcggctcgg	caatcaatca	ggacggacgg	agcaatggca	tcacggcgcc	gaatctgcag	25620
gcgcagaagg	cggctctgca	agaggcgggtg	gccaacgcgc	acatcgatcc	atcccacgta	25680
tcgttgatcg	aggcgcattg	cacgggcacg	tcgctggggc	atcctatcga	gatcgaggcc	25740
ctgcagtcgg	tctacgacgc	gccggactct	gcgccttgct	tgctgggttc	cgtaaagacc	25800
aacatcgggc	atctggaggg	cgcggcgagg	atcgccgggc	tgatcaaagc	cgtactcgcc	25860
ctgcagcatc	gcaccattcc	tccgcacctg	cattttcgcc	ggctgaatcc	gaacatctca	25920
ctggacggca	gccggtttcg	catcgccacg	gaatcgctgc	cgtggacgtc	ggaaggacgg	25980
ccgcgtctgg	ccggcgtcag	ctcgttcggg	tttggaggga	gcaacgcgca	cgtcatcctc	26040
gaagaggcgc	ctgcactccc	tttgccgaag	ccggtcacac	gcccgcagct	tctcactctg	26100
tcggcgcgca	ccgacgaagc	gctcggcgaa	ctggccggcc	acttcgcgga	gttccctgcag	26160
tcgcacccga	atgcgttgct	gtccgacgtt	tgcttcacca	gtcagggttg	gcgcgacgca	26220
tatagtcacc	gcttggcgat	caccgcgcga	gatgcggcag	aggctgtagc	ggcattggcc	26280
gcggcgccgc	ggcgcgaagt	atcgttgcgc	cggcgccggg	caatcgcttt	tctcttcacc	26340
ggccagggcg	cgcagtacgc	cggcatgggc	gcagagcttt	ataaaacgca	gcctgttttt	26400
cgcgacgcgc	tcgatcggtg	cgcgatttgg	ctccgtcccc	agctcgatgt	tccgctgacc	26460
gttctcttgt	tcgagtcggg	ttcgccgttg	cacgagacgg	cgtataccca	gccggcaatg	26520
tttgcccttg	aatgggctct	ggctcagttc	tggctgtcgc	tcggcgctccg	gccggactac	26580
gtgctggggc	acagtctcgg	cgagtatggt	gcggcggtgtg	tggccggcgc	ctttagcgtg	26640

gaggacggcc tgcggctggt gaccgccagg gggcggtctg tcaatgcgt tccccgcggc 26700
 aaagcggta tgcgtcacgc caatccgagc cgcacgcggc cgctcgccgc caaggtggca 26760
 gtcgccgcat cgaatgcgcc ggaccgcacc gtgatctccg gcacggctgc agaaatcgcg 26820
 gaagcgcaag atgacctgca tcgcgccggc gtggaaacgc gagagctgaa cgtatcgcat 26880
 gcgttccatt cgccgctgat ggatccgatt ttggacaagt tcgaagcgct tgcaggtgcg 26940
 atcgcgatc agccgctggc gatcccgctg gtgtcgaacg tcagcggagc cgtattgccg 27000
 aaaggcacga cactcgacgc ccgctactgg cggcgacagt tgcgcgaaac cgtgcagttt 27060
 gaaagcgca tgcgaacct ggccgaccgc gagtgcagc tgtttctgga aatcgggccc 27120
 catcccacgc tcaccacgct gggcgcatat tgtctgccc atgacggcgc ggtctggctg 27180
 cactccctat ctaagggaag atcggattgg tccgtgctgc tggaaagtct tggcgccctg 27240
 tttaccgcgg gcgtgaatcc cgactggcgc ggtctctatg ccggggaatc acccagccgc 27300
 gtcgcgctgc cgacgtatcc gtttcagcgt gacacctca gcctgagacg cgtaccgcgc 27360
 agagagccgg cgcgcggcgg catggtggga gcgcgcctca acagcgcgtt gggcgatgtc 27420
 atcttcgaaa attcgctaac cacggagacg cctctgctcc atgagcacgt gatctacgac 27480
 gcggtcattg tgcgcggcgc ctggcacgtg tcggcatttc tcgaagcggc acaggaagt 27540
 ttccggtccg ttccttgccg cgtctccgat gtcgatgac ggccagcact ggccatccc 27600
 ccgatacgc cggtcacggg gcaagcgatt gtcacaccgc gcgaggacgg cgaagcaaag 27660
 gtgcaggtct tcagccagga tggcgattcg tggaaagtcc acacggcagc cagtctgcgc 27720
 gcggcgactg ccggcgccgt tcatttcgag ctgcccggcgc agccttccga agtcatttcc 27780
 ggcgatgcgt tctacggcgc gatgaacgca cgcggcgctc atcttggccc cgccttcagt 27840
 tgggtggaag aagtcggcg tcgcgatggc gaggcgctgg ggcaatgag tctgcccgtg 27900
 gctgaggatg gcgcgaacgc ttaccggctg caccgcggcc tgatcgattc ttgttttcaa 27960
 gtattcgag cgacttggcc cgcggagcgt tgccagccc gcgcatacgt gccggtcggg 28020
 atcgaagcgg tgcgcttcta ccgtccgccg gcaggttctc tgcgctgtca tgcgcgtctg 28080
 cgcccagact cgagcggccc gttegtcggg gatctgacgc tggttgaaga gaccggcgcg 28140
 gtcacgcgc agttttccgg actggctgta atgcatgcc gtacgtgca atccgcacag 28200
 tcgtggctgc aggatgtgca gtggcaggag tgcgagcgat cgacaacgtt gaagtccgac 28260
 ggccctggca agccggagga ctggttgctg tgtgccggcg cagacgatgt cgcgggtt 28320
 atgccgaag agctgcgcgt cgtgtccggc gtcactctcc gccaggcgct ggaacagacc 28380
 cagactttgg tcggccgccc ggccgggctc tggctgatca cgcgcggcgt gcacgcac 28440
 agtgatgac atgcgactcc cgtcgatcct ttccaggctc cactgtgggg actcgggcag 28500
 gcgatcgcc gcgagcatcc cgagctgtgg ggccggctga tcgacctcgg ttgcgacaat 28560
 gccgacatcg ccgccgccat gctgctggat gaaatccgtt atgccggcga cgacaaagcg 28620
 atcgatttgc gcaacggacg ccgctacgtt cgcggctgg tgcggcaca ggaaacgtcg 28680
 aagcggccgc ctgccatttc agccgacggc gtctatctga tcaccggcgg tctcggcgca 28740
 ttaggacgaa ggggtggcag ccgcttgatc gagcaaggcg cgcgcgcgtc ggtactggtc 28800
 ggccggcata cggaggcagt tgccgatctc gagcaactcg gggctgcagt catggttgct 28860
 gcttgcatg tgagttccga gcaacagctg gcggcgctgc tggcggaccc gcgcaccag 28920
 ccgtgcgtg gagtgcgtgca tgccgcaggc gtgctcgatg acggggtagt tacagaacag 28980
 acgtgggctc gtttcgagaa ggtgctggcg ccgaagctgc aggggtgctg gaatcttcac 29040
 cagctcactc gccaccatgc gctcgacttt ttctgactct tctcttccgc cgttcgctg 29100
 ctcggttccg ccggacagag caattactcg gcggccaacg catcttctga cagccttgcc 29160
 cacatgcgcc gcgcgcaagg actaccggcg ctgagcatca attggggacc atggcgggc 29220
 gaaggcatgg ccgcgcgcat cgcgcggcaa ggcctgccgg gggtaaccgt gctgccggc 29280
 gaagtgggtg cgcgcactct cggcgatctg ctggcgagaa ctgccgctca gatcgcggtg 29340
 ttccaagtct ccgccgaaaa aaggcggagc ccggcgagcg atcccggtt catccagcaa 29400
 ctaccgaag ctgcgcggga gcggcggcag gaactgctgc agatgcgcat ccgcaagcag 29460
 gccggcgggc tgctggcgct cgatgcgtcc aagacgctcg acccgcgccg gccgctcaag 29520
 gaatacggac tcgattcgct gatggcgctg gatctggcgc gcgccatcgg agagctggtg 29580
 cgcaagagcc ttcgcgcgac attgctatac gaccatccga ccgtcgagaa attggccggc 29640

catgtcctcc	gcgaactcgg	actcgacgtc	cccagcgatt	ccctcgtcga	tgaagtgcgg	29700
cagctgtccg	agcaggagat	ggcggcgttc	atcacggaaa	ccttgaccca	tctgggagag	29760
gaacgatgag	cgatctcact	cctcttcaac	aggcggtcct	ggcgctcaag	cgcacgcgag	29820
cgcgtctcga	cgaactggag	agcgtccaca	acgaacccat	cgcgatcgtc	ggcatggctt	29880
gccgttttcc	cggcgcggaac	tcgccggaag	cattttggca	gctcctgcac	gatggcatcg	29940
atgccatccg	cgaatttcc	gcgggcccgtt	gggatgccga	tgcgttttac	gatccccgatc	30000
ccaacgcgcc	gggaaagaag	tacacgcgtc	tgggcggatt	cctcgatggt	gccgtcgacg	30060
gcttcgacgc	cggctttcttc	ggaatcacgc	cgcgcgaggt	cgcgggtctg	gatccgcagc	30120
agcgctgct	gctcgaggtg	gcatgggaag	ctttggagcg	tgccgggtcgg	ccgcccgcaca	30180
gtctcgccgg	cagcgacacc	ggagtgttca	tcgggatcag	caccgacgac	tacagccggc	30240
tgaaacctac	cgatccggcg	ctcattgacg	cctataccgg	taccggaacc	gcgttcagca	30300
ctgccgcggg	acggatctcc	tatctgctgg	ggttgcaggg	accgaacttc	cccgtcgaca	30360
cggcgtgctc	ttcctcactc	gtggcgggttc	atctggcgtg	ccgcagcttg	cagtcgcgag	30420
agtgcagcat	ggcgctggcc	ggcgcggtga	acctgattct	ggcgccggaa	agcacgatct	30480
acttctgccg	cctgcggggcc	atggcgcccg	atggccgttg	caaaagtctc	gctgcctccg	30540
ccgacggtta	cggccgcggc	gagggatgcg	gaatgctggt	gctgaagcgg	ctgtccgatg	30600
cgcgcgtga	cggcgatcgt	attctggcgc	tgattcgcg	atcgcccgtc	aaccacggcg	30660
gccgcagcaa	cggcctcacg	gcgcggaacg	gtccggcgca	ggaagccgtg	attcggggcg	30720
cgtcaagaa	cgcggcgatg	gccccgcgg	atgtcgatta	cgtggaagcc	cacggaaccg	30780
ggacgcgcgt	gggagatccc	atcgaaactgc	ggcgatggc	agcgggtgctg	ggcgaggggc	30840
gtgccgtcga	ttctccgttg	atcgtcgggt	cggtgaaaac	caacttcggc	cacctggagg	30900
cggcggcagg	tatcgccggc	ctgatcaaga	ccattctcgc	cctgcagcac	cgagagattc	30960
cgcctcatct	gcatttcaac	gcgcccaccc	cgcacgtact	ctggaatgag	ctgccgctaa	31020
agatagccac	cgcattgttcg	ccatggccct	ccaacggccg	cccccgagtt	gccgggggtga	31080
gctcgttcgg	aatcagtggc	accaattcgc	acgtcgtcct	cgcagaagcg	aagacgaatg	31140
tagaagcgaa	gacgaatgta	gaggcggaaga	cgaatgtaga	ggcgaagacg	agtgaagagg	31200
tcaaggcgag	tgtagaggcc	aaagggaatg	tggaggctaa	ggctagtgtc	agtgtcccc	31260
tctcagagg	ggacagccgc	ccgcgaagcg	gcggcggggg	gtcgggcccgg	ccgccagcc	31320
gcgagggaagt	gccgggtccc	gatcaactcc	atgccgaaga	cggccgcgaa	tacctcctac	31380
cgttttcggc	gcgccatccg	caggtctctgc	gcgatctcgc	cggcgccctat	cgcgatgggc	31440
gcttttcacgc	tccgctctcc	gcgtgtgtt	ccgccgccag	cctgacgcgc	agtcactacg	31500
aacatcgccg	agcgtttgtg	gcctcatccc	tgcccagagtt	caatcaattg	ctcgaggcct	31560
tccggcgcaa	tgaaaccaat	cgcggcgctc	ccaccggttt	cgcgatccc	ggagtctctc	31620
cgaactcgc	cttcattctt	tccggccagg	gcggacagta	ccgcgcgatg	gcgtatcgcc	31680
tgtattccga	cgagcctgtc	ttccgatcgg	cgatcgaacg	ttgcgacgcc	gccttccgca	31740
gcttcgtgga	atggcggtt	gcggacctgc	tcgccgacga	gtcgggagca	tggctgagcc	31800
agatcgatcg	cgtgcagcct	gcgtgttctg	ccgttcaa	cgcgctggtc	gaactgctgc	31860
aatcctgggg	aattcgccc	gacggcggtg	ccggacacag	catgggagaa	gtggcgggcg	31920
cccatgtcgc	aggcattctc	accctggagg	acgcggcccg	catcatctgt	cgcgcagcc	31980
ggctgttgct	cggacttcgc	ggccggggag	cgatggctct	ggtcgaactg	ccgctcgatc	32040
gggcgaaggc	cgtgctcgct	gaacgcggtc	tcactactgt	ttctgtcgcg	gccagcaacg	32100
gaccacgcag	cacgggtgtt	tcgggagacc	gtgtggctct	cgagcatttg	aaggacgact	32160
tcgagaggcg	cggcgtcttc	tgccggctga	ttcagggtgga	tgctcgcttca	cacagctcgc	32220
aggtggaccc	gctcgagaac	gaattgcgcc	aggaactcgg	ccgcgttatt	gcaaaacgtt	32280
ccgcctgtgc	gttcttctcc	acggttgaag	gacagttgag	cacgggagag	gcgtgcgacg	32340
cgtcgtactg	ggtagccaat	ctgcgacagc	cagtcctgtt	ctgggagtcg	ttgcaggcga	32400
tggctggtga	tgagttcacg	cagttcctgg	agatcagtc	gcacctctgt	ctgacgccgt	32460
cgatcgagga	tagtctgcgg	acgctcggca	taaacggact	ggttcgcccc	gtactgcgcc	32520
gcgacgaacc	ggagcggcgt	gagctgctcg	agttgctcgc	cgcgctctac	gtgaatgggc	32580
agcgtccgga	ctggcgcgcg	ctcgttctgt	ctcccagacac	gcgcctggat	ctgccgacgt	32640

atccctggca	gcgcgagcgc	ttctggttcg	cgacctcgac	gcggcggaagt	ttgccggcag	32700
ttggcggtca	tccgctgctc	ggtcgcaagg	tcgagattgc	gctggcgccg	gacacacacg	32760
tctgggagtc	cgtgctctct	ctggatgcgc	tgccgtttct	cgccgatcac	cggtcaacg	32820
agcttgtggt	gcttcccggg	gccgcttatg	tggagatggc	gctggcgca	gccaaggaag	32880
tggtcgcggg	tggctgcagc	ctggaagaga	tccggtttga	acaaatgctg	gttggtccct	32940
ccgcgggcgc	ctcgcgagtg	caggtcatac	tcgagggaca	cgcattccgc	atctccagtc	33000
tggccgaagg	cggttccgat	tggaccgagc	acgcgcgcgg	caccatggct	gcggcgcccg	33060
acaaggctgc	gcccacgggtg	agcctgcccc	cacttgggga	tcgcatcgag	ggcgatgact	33120
tctatgcggc	cttcgcacgc	caggggatgc	attacggcga	caccttccgc	ggcatcgccg	33180
aagtgtggcg	gcgcgacggc	gaggcagtg	cgcgactgag	cgtgccggat	gccgttcgcg	33240
aagcagagtc	cggttacacg	cttcacacctg	ccttgctcga	tgccgtgttg	caggtgctgg	33300
gcgcgacgct	tggcggcgaa	ggcagcgccg	gtccttgctg	gcctgtcgcc	atcgaacggg	33360
tgactgttt	cggcagaccc	gccggcgatc	ttaggggtgca	tgccgggctg	acggggcgcc	33420
tcgagggcga	tgtcaccttg	tgtgatgcgg	aaggccacgt	catcctcgag	gtccaaggcc	33480
tgctgtccca	ggaactggag	cgccaatccg	aatggttcca	cgctatggaa	tgggagccgc	33540
agctgctggc	cgagagtcca	acggcaacgg	tgtcgggtgc	atggctggtc	attgccgatg	33600
ccggcgccat	cgcagcccg	gtggcgcgag	ggctgggcac	aaacacgggt	gtgatttcgg	33660
gtcgcgatgc	cgagataccg	gatcagcctt	accggggcgt	cattcactgc	gggagcctgg	33720
atgagaccga	ggatgagacc	gatccgtcgg	ctgcgggggg	aaccgcctgc	gaagacattt	33780
tgcgcatcgt	tcaagaattc	ggagtgggac	gcatacagct	gacgaaacaa	gcgtccgacg	33840
ccgaatcgca	gcatccgcga	atctggctga	ttacggcggg	cgttcatgcg	gagcatctgc	33900
agatgccggg	ggtgcccgcg	cgggcaccgg	tgtggggctc	gggacgtacc	atcgcgcccg	33960
agcatcccg	gttcgcttgc	acctgcacgc	atctcgacac	tgccggtgaa	gtcgaggtgc	34020
aggcgctctg	ccgagagatt	ctcgcgggga	gttctgaacg	tcagggcccg	g	34071

<210> 115

<211> 4615

<212> ADN

<213> bacterie

<400> 115

actgcagtgc	ccggaatcgg	cgggtggactt	acagcagccg	ctggtgcgta	tgggattgga	60
ctcgtcatg	gcggtgcaat	tacgcaaccg	gatcgatacg	gatctgcgcg	tcttgctgcc	120
catggtccga	tttctagacg	gccccagcgt	tgcggaactg	gccagggatc	taagcgatct	180
aagcggcctc	agcgaacgca	cgacggtggc	gccggaacct	gcggcgcagg	cctcggttcc	240
tgccctctcc	tacctctca	gcgcgggcca	gcaggcgctt	tggtttattt	accgaagcgc	300
gccggaaagt	cccgcataca	acatcgcggtg	gatcgcgcgc	gcgagaggcg	ctttcgatcc	360
gcaggcggtt	cgccgttcgc	tgcaggacct	ggtggatcgt	catccggcgc	tgcgaaacgac	420
gattgcggag	agtggcgggc	caccggttca	aacgggtccac	agcagcgctc	cgggtggattt	480
cgaagtgatc	ccgtgttcgc	cggacgatga	ggcgggtgctg	atcgacggcg	tcttccacgc	540
gcccttcaat	ctcggcgaaa	actgtttccg	ctcgcgtctc	ctggtgcagt	cggggaagga	600
tcaggttctg	gccatcggtg	tgcatacacat	cctcgccgac	ttctggtcac	tgctggtgat	660
ggtggatgaa	ctccgcagta	tctacctcgc	gaggacagct	ggcgggtccgc	ctgtcgcgcc	720
gccggtcgcg	agcttcgccc	ctttcggtccg	ctggcagaaac	gaactgttgg	ccggaaccga	780
gggcgagcgg	ctttggaaact	actggtcctc	gcagctttcc	ggccagcttc	cggttctgaa	840
tctcccgtcg	gatcggtcca	gtccgcccgg	gcagagtttc	cggggaaaact	ctcactcggt	900
ccgaatcgaa	cccgcgctga	ctgcgaaact	gaaggcgctc	gcgcggcggc	agaacgcgac	960
gctgcgatgc	acgctgatgg	cggcggttca	agtgttctc	tcccgttggg	cctcacaaga	1020
agagatcctg	accggcaccc	tcaccaacgg	tcggacgcaa	ccggaattcg	ccgatctcgt	1080

cggataacttc	gtgaatcccg	taatcctgcg	aggagaactt	tcaggcgatc	cggatttcaa	1140
tacggtgctc	gcccggattc	ggcaaacgct	tctcggcgcg	atcgagcacc	aggagtaccc	1200
gtatgcccgg	atcgtggagc	ggttgggtcc	cggactgcgg	gttctattcg	tgctccagca	1260
gcctcatcgc	attcccgaat	ccgtgccgtt	catgttgggt	cagtcggcg	gtcgcattggc	1320
ctggggcagc	ctcacactgg	agtccctggc	gatgccgctg	cgacagagcc	ggtttgacct	1380
ggatctgatg	atggtcgaaa	ccgatggagg	cctctccgcc	tttctgcaat	acaacacgga	1440
catttttgat	gctgccacga	ttgaacgtct	ctccttgcac	ttcgccgtgc	tgctggaagg	1500
aatcgcgag	aatcccgcct	gtccagttgt	cgatctaccg	ctgctgacaa	cccgggaacg	1560
catccagctg	ctcgaagagt	ggaatgcgac	cgccgcggaa	ttcccgtccc	aatgcgtgca	1620
cgagctgttc	gaagctcagg	tggagttagc	gcccgcagcc	atcgcggtga	gcttcgggtga	1680
gcagaatctg	acatatcgcg	aactcaacgg	gagcgccaac	cggatcgcg	actatctccg	1740
ctcgcgcgcc	gctggaccgg	gcgaaatggt	tggcatccat	gtcacgcggt	cgctcgaaac	1800
cgtcgcaggg	ctggtggggc	tcctgaaggc	cggcgcgccc	tacgttccgc	tggaaaccgga	1860
atatccggcg	caacgtcttc	ggctgatgct	ggaagagacc	aggccggtcg	ttgtgctgaa	1920
tgtcacggaa	tcggaagtat	ggacgcagcc	cgacaccaat	ccgaaccgcg	tcgcgactcc	1980
cggcgatctc	gcctatgtcc	tgtacacctc	cggttcgacc	ggccggccga	aaggcggtgca	2040
aatcacacac	caggccgtcg	tcaattttct	ttcgtcgatg	cggcatgagc	cgggcatcag	2100
cgaccgcgat	acgtctctcg	ccctcacgac	gttcattgtc	gacatttccg	cgctcgagat	2160
ctttttgccc	ttgagcgccg	gcgcgcgcgt	cgtggtggcg	aaccaggaga	cggccgtcga	2220
tggtagaggg	ctggcgaggg	aactcgcgcg	cagcaaagcg	acaatgatgc	aggcaactcc	2280
cgccacctgg	cgtctgctgc	tcgcatccgg	ctggcccggc	gaccgcccgc	tgacggcgct	2340
ctgcggcggt	gaagcccttc	ctcgcgatct	tgccgaccgg	ctcctgcaac	gaaccgcggc	2400
gctatggaat	ctttacggac	ctaccgaaac	gacaatttgg	tcgcccatcc	aacgggtgac	2460
gacaggtgac	ggaccggttt	cgattggcgg	ccccatcgca	aacactcagc	tctatgtgct	2520
tgacgatcgg	atgcagcccc	caccatccgg	tgttgccggc	gaactgtaca	tcggcgggcg	2580
cgggctcgcc	cgtggatacc	tgaatcgctc	ggaactcagc	gcggacaagt	tcgtcgccaa	2640
ttcgttcgac	cctcatggca	ctcggctgta	tcgcacggga	gatctcgccc	gccgccaacg	2700
cgacggcgcg	ctcgagtatc	tcggccggat	cgaccaccag	gtgaagatac	gcgggttccg	2760
catcgaaacc	ggcgagatcg	aggccgcggg	ccgcagtcac	ccggcggtcc	gacatgctgt	2820
ggtcaccgcc	agagaaaatg	acgcggccgg	taagtatctg	gcggcctaca	ttgtccccct	2880
tgctgacggg	catcgcgcca	cggcagccgc	cgacacattc	cacgaccgag	tcgagtccga	2940
gcacgtgacg	cagtggcaat	ccgtctggga	caccacatat	gaacagaatg	cgccgaacgc	3000
ggatccggag	ttcaacatcg	tcggctggag	aagcagtgtt	accggagagc	cgattccagc	3060
tgcgagatg	cgggagtggg	tgcaggattc	cgtcgatcgc	atcctggcct	cgcggcccg	3120
tcgcgtgctc	gagattggct	gtggtacggg	actgctgctc	ttccgcgtcg	ctccccactg	3180
ttcggagtac	tgggccacgg	acttttcgca	gaaggcgctg	gactacatcg	ccgctcacgc	3240
ggaccgcacc	ggcctggcaa	atgtccgcac	gttcgggcag	gcggccgacg	acgcgtgcca	3300
gatcgacagt	cgctcgtgcg	atgcggttgt	tctgaactcc	gttatccagt	acttccccgg	3360
cgaagcgtat	ctgcggcgcg	tgctggccga	ggcggtgctg	gtggtcaaac	cgggcggcat	3420
cgtatttgct	ggcgatgtcc	gcagtctccc	gctgctggag	acgtttttacg	cttcttttaga	3480
agttcagcgc	gcacccgcgt	cgttgaccgg	gaatgagttt	cggcaacgcg	tgcgttcgct	3540
cgcgtcgag	gaagaggaac	tcgtggtcga	tcccgcgttc	ttcttttgctc	tccgcgaaca	3600
gattccggag	atcgggccgga	ttgaaatcct	gccgcgtcgc	ggccgggtcgc	ataacgagct	3660
gaccgccttc	cgctaccagg	cgatcctgca	tatcggatcg	cgggaagcgg	aggagccgga	3720
atcggatcgc	aggcgttgcc	agaccgcggc	cgaaatacgc	agagtactga	cggacgctca	3780
gccggagtgg	gccgcattta	ccgagattcc	gaacgcacgg	ttgaccgccc	aaagcgccat	3840
tgtgacctgg	atgaacgggtg	acgaagctcc	agagacactc	ggggagtgtg	gggaccgggt	3900
gcgccagacg	tcgccttccg	gcgtcgatcc	cgccgatcta	tggcgatatg	acgaagacct	3960
gccgtaccgc	gtggcaatcg	actggagcag	tcatgggcca	cacggacgct	tcgacgcgac	4020
cttctgccgt	gcggcgggccg	gtccgcgggc	ttcccgtccg	cgacgccgcc	tggccggccc	4080

```

gtatacgaac gatccgctgc gagccgtcta tacgcgcacg gttgtgcgc agttgcgtac 4140
tcatctgaag gagaagctgc ccgactacat gatccccgacc gcgtgggtcg tgctccacga 4200
aatgccgctg acgccccaacg gaaaaatcga ccgtaacgcc ctgccccgatc ccgagccccag 4260
ccggcgagcc cagcgccgaag cattcacgcc tccggaaact ccggtggaac aggtactcgc 4320
ccacatttgg ggcgaggtgc tcggcatgga tggcatcggc gtccatgatc acttcttcga 4380
ctctggagga cattcgctgc tggtcacgca gatgatcgcc cgcggtgcgc acatgctcca 4440
cgtggaagtg ccttttcgaa ccgtgtttaa cgcgccacg gttcgaggct tcgccgtcgc 4500
tattcaggac ggcgtagacc caggatgggc aaggcgagcc gccgatttgc tgatcgctgt 4560
tcccaaagt tcagatgttc aaatcgagcg tatgatgagc gccgcccaag actag 4615

```

<210> 116

<211> 8301

<212> ADN

<213> bacterie

<400> 116

```

atgcagaatt cgtcgccaaa taccatagac ctctcgctcg cccgccgcca attgctcgac 60
cgtctgctgc aggaaaacag ccccgaaacat cgcacccgc ggcgtagaaa ccgggatgcc 120
gcacccttgt cgttgccccg gcagcggcctt tggtttctcc atcagctcga ccgggattct 180
cccgccctaca acattcccat agcgtcgcac atccgaggtc cgttgatat tcgctcctc 240
ctgcggagtc tggaggccgt ggtgcagcgg cagcagagcc tgcgcagctg cattggcggg 300
gtggatggag aggcgcgcca gagcctcctg gcgcgagtga cactggaact tccggttgtt 360
caggctgacg gaatcgcaga agcgcggcaa atggccttgc gtgatgccc gatcccgctc 420
gacctgcgaa aacccccgct tctgcggacc aagctgatct gcctcgatga caagcagcag 480
attctcctgc tgacgttgag ccacatcacc gcggatgcgt ggtcggtcga gacgttcgtc 540
cgcgacctga cgcgatcgta cgaagcgttc gtgcaggggc ggccatcgcc gctcatggaa 600
ctgccgattc agtatggcga ctgggcccgc catcagcaga cgtcgctgaa ccaaaccgcg 660
cagcagtact ggaagaaaca gctgtcgggc acctgcctt tcctcgacct tcctaccgat 720
cgcccccgcc ccgcgcagca gacctggcgg ggcccggtgg agaccacagc cctcgggcgt 780
gatttgaccg atggactcca cgcgtttgcc ttgcgtgaag gagcgacggg gttcatgacg 840
gcaatcgcg cgtttcaggt gctgctgcac cgtatatacc cgcaggaaga catccttacc 900
ggggttccag tcgcggggccg tacacaacga gaaacggaag gtctcgtcgg ttgtttcgcc 960
aacatgatcg tctgcgcgg cgatctgcgc gacgatccgt cgtttcgag tcttctcgcc 1020
cgcacccgcg acaccgcttt gagcgccctc tctcatcagg actttccttt cgaacgcctg 1080
gttgaggaa tcgcatcctc gcgggacctg agccggctgc ctgtatttca ggtctccttc 1140
gcgctgctgc ccgatgcgcc ggccatcacc gtcatgcctg ggctcaccat ctgcgcgag 1200
tacatgcaca acggcggatc caaactcgac ctcggcgtga ccctcgagcc atccggcgat 1260
ggactgatgg cgtccgccga atacaacacc gatttggtcg atgcggcaac catcgccctc 1320
ctgctcgatg cgtaccgaac cctgctggcg agcgtggtga cggatcccga cgtccgcatt 1380
tcaaccgctg cgtgttgtc ccccgcggtc cgaagccgga tgctcgagca gcacaatgcg 1440
acacggcgcg atgccggtcc gaacgggtgt gcgcataaac tggtcgaagc tcaggcggaa 1500
cgactccgc acgcgctgc cgttgtcttc gaagaccatc agttgacctc cgccgagctg 1560
aatgcgcggg ccaaccgcct ggctcatcgt ctgagcgcat ccggcgcggg cccgggaaag 1620
atcatcgctc tggcgatgga gcgctcgctg gagatggtga ttgcgctgct tgcgattctg 1680
aagtccggca gcgcgtacct gcctctcgat ccgcgcacc ccaaggatcg tctcgcccg 1740
attctcgatg aagtgcgaacc gcacgcggtc ctcacgcagg aggcgggtgg tgagatgatg 1800
gcgatgatgg cgatgatggc ggtcgccgtc gaaccagaag ctgcgaatct cgtcagcggc 1860
agcaagccc acgatctcgc ctacatcata tatacctccg gatcgacggg gcgaccgaag 1920
ggcgtggaga tccgccactc gtcgctagtc aatctgctgc gctccatgca gcgcgagccg 1980

```

ggtctgacag ccgccgatgg gctggtcgcc gtcaccaccg tgctattcga tattgccgga 2040
 ctggagatct ggctgcccgtt gatcaccggc gcccgcgtea tegtccccac ccgcgagatc 2100
 gtggttgacg gcgagcggct caccaccctg ctggataagt cgggcgctac ggcatgcag 2160
 gcgaccccga gcggttggcg gcaattgctg gattcgggct ggaagccggg taaaggcttc 2220
 cgtgttttct gcggcggtga agctctgccg ccggaactgg cgcgccgat tctcgatagt 2280
 ggcgtagagc tgtggaatct ttacggaccg acggagacca ccatatggte ggccgtgcac 2340
 aagacacaaa gactgggtgc ctccgatagc atcgtgccga tcggccatcc catcgacaac 2400
 acgcagttat acatcctgga ttccgcgatg gagccgggtc ccccgaggat tccgggagag 2460
 ctgtacatcg gaggagcggg actggcgcgg ggctatcatc gcaaccccga gctcacgcgt 2520
 gagaaattcc gcgagtggcg tgatcgagga cgcatttact ctaccggcga tctggctcgc 2580
 taccgttccg acggcgcgagt cgagtgcctg ggacgagtcg atcgccagat caagctgcgc 2640
 ggggttcgca tcgaaccggc cgagattgag gcccgcatcg agacgcacat tgccgtgaag 2700
 caggcgatta cggctcgtgaa ggacgatcgg ctgatcgccat atctcgttcc ggcaacgggc 2760
 gacgtgcgcg atctgcagag cgatttgccg tcgtggctgg caacgcgcct tcccgattac 2820
 atgatccctt cggcgtttgt cagcctgtcc tcccttcgcg tgacgcccga cggcaaaatc 2880
 gacgcgaacg cgcttcccgg ttgcccaca acgcccgttg ctgctcgcga gccgatgcgc 2940
 ggcgatgtgg tggagacgat tgcgtccatc tggcgtaag ttctgcgcgt ggagcacgtc 3000
 gactatcggc agaacttctt tgatgtcggc gggcactcgc taatgtcac acgggtgcgc 3060
 ggactgctcg aggagcgcc tgggttgacg ctctccgtcg tcgatctgtt ccggcatacg 3120
 acgatcgagt cgcttgccgg cctggcagaa aaatccgaac ccgccgctgc ggaacctgcg 3180
 gctgcggtcg cagaagatcg gatcgagtt atccggatgg ccggccgggtt cccggggggc 3240
 cgcaatgtgg aggagtcttg gcgcaatctg ccgcacgggtg tggattccat cgcaggtctt 3300
 tcgcccgaag atctgctggc gggcggcatc agcccggagg tcttcagga cccgagctac 3360
 gtgccggcca aggtctgct ggacggcatc gatttttctg atgccgcgtt ctccggtac 3420
 agtccgcgcg aagcggagat catggaccgg cagcatcgcg tgtttctcga gtgcgcgtgg 3480
 gaagcgatgg agaacgcggg atatgcggcg cgaagctata agggttcgat cggcgttttc 3540
 gcgggatgcg gcgtcaatac ctacctgctg aacaacctcg ccaccggga gccgttcgat 3600
 ttctcacgcc cctccgcgta ccagctgctg acggccaacg acaaggattt cctggccacg 3660
 cgtgtctctt acaagctgaa cctccgcggg ccagcctga cggttcagac ggcggtgctcc 3720
 acctcgctgg tgtcgggtgg gatggcatgc gagagcttgc agcgcggcgc ctccgacatt 3780
 gccttggccg ggggagttgc catcaatgtt ccgcagtcgg tggggtacct gcaccagccg 3840
 ggcgatgacc tgtgcccgga cgggcgctgc cgcgccttcg atgagtccgc tcaaggcacg 3900
 gtgccgggca acggcgcggg tgtggtcgtc ctcaagcgtc tgagccgcgc tctggccgat 3960
 ggcgacacga tctacgccgt cattcgcgga gcggctatta ataatgatgg cgcgagcgc 4020
 atggggttta ccgctccagg tgtggacggt cagacgcgat tgattcggcg cactcaagag 4080
 atggcgggcg tgaagccgga gtccatcggc tacatcgagg ccacggaac agccacgccg 4140
 ctccgcatc cggtgagat cgcgccatc gctgccaaact ttccgaaaaa cgggaagcggc 4200
 gatgtgtata tcggatccgt caagaccaac atcggtcac tagacgtcgc ggccggtgtg 4260
 gccgggctga tcaagacggg gcttgccgtc catcgcggcc agattcctcc cagcctgaat 4320
 ttccagcgtc cgaatccgcg aattgatttc gcaaactc cgtttcgtgt gactacgcgg 4380
 ctgctcgact ggcccgccgg aaagaccccg agacgagcgg cagtcagttc gttcgggatc 4440
 ggccgaccca acgctcacgt gattctggag caagcgcgc cggtgacgcc ggccgcagct 4500
 gcgcccgaac gatccgcaca tgtgctttgc ctgtccgcca atacagacgc ggccctcgaa 4560
 gaactggtgc gctcgatcg cggccatatg gacgagcgt cacttccgc accgtatctg cattgtggcc 4680
 gtcgattca cggccaatgc agggcgcggt ccaacgactg acggaggcac gacgggttcg catcgcccag 4740
 cggctcgagc agaggtcgt tcttttcacc gggcaagggt cgcaatacgc gggcatgggc 4800
 cgccagttct acgagtcgca gccggtgttt cgcgcgcgca tggatgaatg cgcagctctg 4860
 ctgaatggac ggctcgatct gccggcgctg ttggccgatg acgcgttgct cgacgcgacc 4920
 gccggcgcgc agcccgcgct gtttgctttg cagtgggcct tggcgagtt gtggaagtcc 4980

tggggtgtga cgcccgacct ggtgatggga cacagcgctcg gcgaatacgc ggccggcggtgt 5040
 attgccggcg ccgtcagcct gccgatgctg ctccgcttag ttgccgaacg cggccggctc 5100
 atgcagaacc tgccggaagg tgcgatggct gcggtcagcg ccggcgagca gcgctgtgcc 5160
 gcagcgatca cctcgcgctg ctccattgct gccatcaacg gaccgctga ggtcgtgatt 5220
 tcgggtgcgc cgcaggatat tgagagcgcg ctggcaactc tacgtgcgga gggcatcaaa 5280
 acgcagatgc tggcgttgc gcgcgccttt cacagctcga gcatggatcc gattctggcg 5340
 gacctgcaac gccgggcggc ggcgatcgcg tggcgcaatc ctccgatcgg cttgggttcg 5400
 aacctcacgg gcaaactggc cggcgagggg cagctggcga atccgctgta ctggcgagat 5460
 cacgctcgaa accctgtccg ttccgccgac ggtatccaaa cgctcaagga cgaaggctgc 5520
 gacgtgttcc tcgagatcgg tccaaagccg gttctactcg gcatgggcca aaagtgcctg 5580
 cccgacgacg ccaagcagtg gctgccgtcg ctgcgtaaag gccgcgatga gtgggagacg 5640
 attctcagca gtgtggcgac gctatatcag ggtgggttcg acatcgattg gcaggagttc 5700
 gaccgtccgt attcgcgaag gcgtgtcgcc ctgcggcctc atcctttcga gagacgccgc 5760
 cattggatcg agcggagttc cagaccggaa cctgtagcgg ttgcgagtggt tctcgtcggg 5820
 tgccggctgt cgctaccggg ggcagacggt atcttcgagt cgaaactatc gacggcttcg 5880
 cctctactct cagaccaccg atattacggt tcgggtgggtg ccccgccgtg gtacttctctg 5940
 gccatggcgc tcgaggcgct gccggagggtg tttggcgccg gccggcacac gctggaaaac 6000
 gtgaacttcg cgcacctctc gatcctttca gcggagcgcg acacggctgt tcagctcgtg 6060
 ctttcacaga gcgatgaccg gcatgcctcg ttccgcatac tcagcttgct cgacggctcg 6120
 tggaacttac atgctgccgg caatattgcc gccacgctg gtgtcgctcc cgtgccccga 6180
 ctggctgatg aacgccggcc tgccgtggat ggagacacgt actatcgct gctgcgccac 6240
 ctcgagatag aactggggcc gagctaccgc cgcatacagc gcattcattt cgggtgaacag 6300
 gaagcgctgg ccgcgattga ttccgcaacg ccgctcaatc cccgttggtga attggcggaa 6360
 gccggcctgc aattgcttag cgcgcggcg agtcccgcgc ttgcggatgg ccgcaacat 6420
 ccgatattcg ctccgctcgg tatcgatcgc gtttgtttt acggcagcct ggaggcgcc 6480
 gtatgggggg ccgcgcaaat tctccggcat tcgccggacg gctttaccgg cgaggcgag 6540
 ttgctggact cggagggtcg cgttctcggg gaacttcagg gcgtgagttt ccggcgcgctc 6600
 actcgcgcgt gggcgagcgc ctccggaacg aagcccgaat tgtatgaggt cgagtggcgg 6660
 cccgaaccgc tccgccagcc ttccgcaacg ctacagcctg gggcatggct gatcctggcc 6720
 gacagtggcg gcgcggcccc cgctctggca gatgcgctca cagctcaggg cgagatgtgc 6780
 gttaccgtgc cgcagccgg cgagtacatg tccctagtcg gtgagcgtga ctggcgcggg 6840
 atcgtcaacc tgtacagtct cgatgattat gagctcggct gccgcagcac tctggccctg 6900
 gtgaagtccc tgaagtccgg tccgcggcta tggctggtaa cggccggcgc gcaggcgacc 6960
 agtgccgtgc acaatcccat gcaggccgcg ctctgggggt tcggccgggt gatcgcgcg 7020
 gagcaccggg atctgtgggg cgggctcctc gatctggatc ccgacgatgc gcatgcttcg 7080
 gcggccggcg cggccgcgca gatgcgtgat ttcgacggcg aagatcagtc ggcgtggaga 7140
 agcaaccggc gctacgtgcc gcgactgacc cgccgaccda gcgcgcgagc ggcagtcctg 7200
 ctggtttcgg gcgcgactta ttgatcacc ggcgggctcg gagccctggg acttacagtc 7260
 gcgaaatgga tgggtggagca cggcgccact cgcgtcgtgc tggccgggcg ccggcctcca 7320
 aacgaggagc agcagcgctg gctgcaacag attggtgcga cggcagagac ggtcgacgtc 7380
 agccgggaag aagaggtcgc ggatctcatt cgccgcattc acaccgaaac gtcaccgctg 7440
 cgcggcggtta tccatgccgc ggggtgtgctg gacgacggcg tactgctgaa tcaggactgg 7500
 acgcggatcg caagcgtcat ggcccggaag gcggaaggcg ctgtacacct ccatcatcac 7560
 acccgcgatc tgccgctcga cttcttcgtg ctcttttcat cggcatcctc gctcttaggt 7620
 cctgccgggc aggcaggcta cgcgcggcc aacgccgttc tcgatgcgct ggcgcattac 7680
 cggcgcgagc tgggtttgcc ggcgaccagc attaactggg ggcgctggtc gggagccgga 7740
 atggccgcgc gcaccagcca gtcgatggcc ggcgtggcga gcctctccgt ggacgagggg 7800
 ctacacattc tcgaggccgt cctgcatgaa tgccccattc agattgccgc gctaccggcg 7860
 ggctcgatta ccggcgaggt gctgcgtccc gccgcgctgc cttcacctca actgcgcacc 7920
 cgcttgaacg aagccacacc ccggcagcgc gaagccatcc tcattgcgca catcaggag 7980

```

tcaactggcgc gctttgtcgg catcgcgact tccacaccgc tcgatccaca gcagcctttg 8040
ggtgaactgg gactcgattc gctaattggcc atagaacttc gcaactcgct ctcccaatca 8100
ctggggcagc ctttgcccgc gagtctgctg ttcgactatc cgtcgctcga tgcgatcgtc 8160
agttacgtgc tccatgcggt atttccaccc gaagcatcac cggtggaagc gccggagttt 8220
gagaacctcg cccgcgaaga actggaagcg ctgctcgatt cgcggctggc gcaggtcgac 8280
cagtggttgg agacgaata a                                     8301

```

<210> 117

<211> 5292

<212> ADN

<213> bacterie

<400> 117

```

atgagcgggt cagacgatct cagcaagctt cgccgcgcgc tgattgcgct cgacaaggtg 60
cagaaacgca tcgaccagct ggagagcgcg cgcagcgagc ccacgcgcct catcggcgcg 120
ggctgccgct tccccggcgc atccaatctc gatgcctatt ggtcgttgct gcgcgagggc 180
cgcagcgcgg tacgtgaagt tccaccgcac cgctgggaca tcgatgccta ctacgatccg 240
gateccggcg cgacggggcg aatgtacacg cggtagggcg gcttcacga tcaggttgac 300
cgttttgacg cccggttctt cggcatcgct ccgcgcgagg cgatcagcct ggatccacag 360
cagcggctgc ttctggaagt cacctgggag gcgatcgaga acgcccggct tccaccgcac 420
cggctggcgg ggagccggac cggcgtcttc atggggatct tttccaacga ttattacaac 480
ctgcaaagtc gcggcgggga tgcgcatac gacgcgtaca ccggcacggg caatacggcc 540
agcgttgccg ccgggcgctc ctcgatcac ctcgggctgc agggcccga catggcgatc 600
gacacggcat gctcgtcac gctggtcgcg gtgcaccttg cctgtcagag cctgcgctca 660
ggtgaaagcg acctcgcgct ggccggcgcg gtcaatctga ttctctcgcc ggatcggacg 720
atctacttct gcaagctgaa ggcgatggca gccgacggtc gctgtaaggc attcgatgcc 780
gcagcagacg gctacgtccg cggtagggcg tgcggtgtgg ttgtgctgaa gcgactctcc 840
gacgcgctgc gcgatcgca tccggtgatg gcggtgattc gcggcacggc aatcaaccag 900
gacggacgca gcaatggact gacggcgccg aacgggcccg cacaggaagc cgtgatccgc 960
caggctgtgg gagacgcgcg cttgcagacg ctggatgtga gctatgtcga ggcgcacgga 1020
accggcacgc cgctgggcga tcccatcgaa gccggagccc ttgcggccgc gctgggagcg 1080
gggcgcacca acggcaacaa gctgaagtc gggtcggtga agaccaactt cggccacctc 1140
gaggcggcag cgggcgtggc cgcactgatc aaggtggcgc tgatgctgca gaacgaagcc 1200
attccgcccc atctgaatct gaccacgccc agcccgcaca tcgattggaa cacgcttccc 1260
ctcgaaatcc cggcacggct cacccttg cgggttgac ccggcgggcg gcgcgtcgcc 1320
ggcatcaact cgttcggctt gagcggtagc aatgcgcacg tgctcatcga gcaggcgccg 1380
caacaggccg cgtccagtac gccgcacccg tacctgcttc cgctatcggc gcgcagtccg 1440
gaggcgctgc gtgatctggc gcgcgcatac cgcgacgtgg tgaacgacaa ccccgccgac 1500
acctgctaca cggcgtgcgc tcgccgcact tcatcgaaac accgcgcggc attcaccggg 1560
acgaacgcgc aggacttgat ggccgggctg gacagttttc tggcgggcaa cccgaaccgc 1620
gataccgcca caggttttgt gccgcgcggc cagaagcgaa aagtcgtttt cgttttgccg 1680
ggacaaggat cgcagtggcc cggcatgggc cgcgacctga tggcttctga accggtgttc 1740
cgtgccgcca tcgaagagtg cggccgcgcc atgcagcctt acgtcgactg gtcgctgacg 1800
caagagttgc aggggcccgt cgaccgcac gacgtgattc aaccggccct gttcgagtc 1860
ggggtcgctt tggccggact gtggcgccat tggggaatcg agccggacgc cgtgatcggc 1920
cacagcatgg gcgaagtcgc ggcagcgcac attgcagggt cgctgactct cgatgaagcc 1980
gctcgggtga tttgcctgcg cagccggatg ctgcgcggag tacgcggcca gggagaaatg 2040
gctgtcggtg aattagcgct ggacgaggcc atcgctgcca tcgccgggcg ctcggatcgg 2100
gtctcgattg ccgccagcaa cagcccgcgc agcaccgtcc tgtcgggcca cagcgagct 2160

```

ctgggcgaac tgctgcgga actggaggcg aaagacgtct tctgccgtcg cgtgaaagtg 2220
 gacattgcct cgcacagcca tctgatggac tccgtgtgcg cggcgttgcc gggcgtgggtg 2280
 ggagcgcttc agccgcggcc ggccgccctt ggcattgact ccaccgtcac cggcgcagcg 2340
 attagcgggtg aagagctggt ttctgcgtac tgggctcgta atcttcgccca acccgtgatg 2400
 ctgtcgacgg cgtcgccgc agccgcggcg ggtggctatg atgtgtttct ggaactgagt 2460
 cccaccccg tgttggtcca gccgatccag gaaacgctcg gagatcgggc agcgattgcc 2520
 gctgctcgt tgccggcgga tgaagacgga aacctcgcac tgcgccggac gctgggagcg 2580
 ctgctgacta acggagtcac tccggactgg tctcgtatct atcccaacgg cggccaaact 2640
 cgccggctgc ccaactatcc ctggcagcgt gagegttatt ggatcgatat ccgtccgccg 2700
 caggctcagat ctccaggttt gccctggccgg cggatcccgt cggcgtgcc ggagatgcag 2760
 ttccagatcca ctgtggaggc gaaagatttc gccgatcacc ggctgcacga tgtgatcgtg 2820
 actccgggag cgtggcacct ggcaatggcg ctgcgcgtg cgcgccagg tctcggcgcc 2880
 gggcctcacc atgtcgaaca cgtgtcattg acgggcgcgc tgacgctgcc ggaaaacgat 2940
 gctgccaggc aggttcaact ggtactccgt catgaagagg gcggcggagc ttccttcgcg 3000
 atctacagcc gcgaggatcc ctggaagctg cacagcgaag gcattgctga ggccggcgat 3060
 tccacggcat ccacgatct ggatgcgatt cgcgcccgct gcacggcgga gctcacagcc 3120
 gatgccttct attcgcgact gtgggatcgc ggctatcact tcggtccac cttccgaacc 3180
 atcggcccca tctggcgcg caacgggtgag gtgctttgtc gcgtggacat tccgctgacg 3240
 gaaatgcaga cgatcgactg ctgtctgcag ttgcccgcg ccctcgtcca tcacgacgat 3300
 ttgaaagatg tgcattgccc ggtaggtctg gaccgattct cgtcgtctga agtgccact 3360
 ggcccggtct ggggatacgc ggtcttgccg ccgattcca cggtggtatg ccgtctcgtc 3420
 accggcaccg gcagcgtggt ggccgaattg gtggggctgc agtcgagagt cggccatagc 3480
 ggccagctcg gcgaatcgga gattcccacc tggacgggtc aatggaccgc gtcggttcgc 3540
 cgcggcgatg ccaatgccgg caatgctggc ggaccttggc tcgtcatcgg cgagccggcg 3600
 attgccgaga ctctgcaaaa gcgcggccaa acctgccga cggccgatac gtgctcgggt 3660
 ccgcgtgcc gtcaaatgt gtactgtccc tcgcgcgca tcgacgacct gctttccgta 3720
 ttgcgcagca tcgtgcaagc gggctggcct gagccgcgc gcctgtggct gctgacgcgc 3780
 ggatctgccg cggttctcaa ctccgacaaa gatattgata ttccgacaagc ctggctgcac 3840
 ggaattgggc ggacgattgc ctatgagcat cccgagctgc gctgcacgct cgtcgatctc 3900
 gatgcgcaca gcaacgactg cgggcatctc gcgacgtga tgcgtgcgaa tatcgagag 3960
 gatcaagttg cgatccggca aggcacggta tggcgccgc gcctcagtct tcacaagatc 4020
 ccacccgcac ccgatgtggc gttccgtgcc gacgcaacct atctgatcac gggcgggctc 4080
 ggccgactcg gactgcaggt ggccggatgg ctgcgcgcg ccggagcgcg ccatctcgtt 4140
 ctgctgggac gcagcgagcg tccctcgcca caactggaag gtgtcaacgt caagatcatc 4200
 catgcggacg tggcggaacc gcagcagcta tcggatgcgc tcgcgatcat cgatcgcgac 4260
 atgccgccgt tgcggggcgt gttccatctg gcaggcacgc tggccgacgg catgctgctc 4320
 aatctcacga ccgaacgctt cgaagccgcc atggctccga aagtagccgg cgcgtggaac 4380
 ctgcacgaac tcaccgccgg ccggccgctg gatcatcttg ttctcttctc ttccgccagc 4440
 gcgacagtgg gatctcccgg ccagggaac tacgccgcg gcaattcatt tctcgacgcg 4500
 ctggtcatc tgcgccgcgc ccagggtctt cccgccgtca gcacgcgtg gggaccgtgg 4560
 acacaggttg gtttgccgc acaggcgaac cgcggagacc gtctggccgc gcgcggcatc 4620
 tcggttatcc aaccgcaaca gggattgcgc gcgctctaca aagcattgac gcagattcgg 4680
 ccgcacgtcg ctgtcatgaa cttcgatatc gcgcagtggc tccgttacta tccgtcggcc 4740
 gcacgatgt ccctgctggc cggcatcgca cccgcggccg cggacaccaa accggcgcc 4800
 gacatgcgca gcgagctcct ggcagttcca gccggggcg agcgcgcgc gcggctggaa 4860
 acgtgctga tgcacgaagc cggacacgtg ctgcgcttcg atccagcgaa actcgacggc 4920
 agagcgacgc tgggtgatct cggattcgat tcgttgatgg ccctcgagtt tcgcaaccgt 4980
 ctggaagccg ggctgcgcgt caagctttct gccaccctga tctggcgta cccgacattc 5040
 tccgccctgg cgcagcatct cgcgcgacaag ctgcgcctgc cgtggaaag catggccggc 5100
 aatgctgaac ctccgaccgt tgctgccgtt gctacccttg ctaccgttg caccgccgcg 5160

ggcgaggacc ggagtccccgc cgctgcagac gatctcgacg ccgtcgcaaa ccagatcgcc 5220
 ggggttggggg acaaagaaat cgaagctttg ttgaaacaga agttcgctca tttttcagga 5280
 gcctccgagt ga 5292

<210> 118

<211> 6462

<212> ADN

<213> bacterie

<400> 118

gtgagttcga tatccgagcg attccccaac cttacgccgt tgcagcaggc gtacctgacg 60
 ctggagcaca tgcagcgacg tctcgatgcg gccgaacgcg acgcgcgcga acccatcgcg 120
 atcgtgggtc tgggctgccg gtttccgggc ggcgatgggc ccgatgagtt ctggcagatg 180
 ttgcgcagtg gaggatgagc tattcgtgag gtaccgcctg gacgatggga cgaggagtcg 240
 gtccggcgca tcctgaaatc gttgaacccc gccacgccgg tgaagattca agccggattt 300
 ctcgattcca tcgatgggtt cgacaacgat tttttcggca tttcgccacg cgaggccgctc 360
 agcattgatc cgcagcagcg gctgctgttg gaagtggcgt gggaggcact ggaggatgcg 420
 gggcagacga tgggaagggt ctccggcagc cgcacgggcg tcttcgctcg gatccacagc 480
 caaagcagcg actatttctg gatgcagacc gccgatggcg cgcgcacga tccgtatacc 540
 gccaccggca cggcgcatag cgtgatcgcc ggccgacttt cctatttgct gaacttgcaa 600
 ggaccagca tcgcgctcga caccggcctg tegtcttcgc tggcggcggt tcatctggcg 660
 tgccagagcc tgcgcagcgg cgagtgtacg ctggccgtgg ccggcggagt gaactctgcgc 720
 ttctcgcggg agtttatgta cgccacctcg aagatgggaa ccgcctcgcc cagcggctgc 780
 tgccgcgcct tcgacgcggc ggcggacggc atcgtgttcg gagaaggctg cggcgtggtg 840
 gtgctgaagc gctgtccga tgcactcgcg gccggagacc ggggtgtggc cgtggtgctc 900
 ggctccgcgg tcaatcagga tggcgcctcg gccgggtca ccgctcccaa tgtcgtgtct 960
 cagcaggtcg tcatccggtc ggcattggcc aatgcgggcg tcgcggcgca gcagatcggg 1020
 tacatcgaag cccatggcac ggggactccg ctccggcgtc ccatcgagat cgaggcgctg 1080
 gcggaacccg tcggcctccc gcgacctgtc ggcgatgtgt gcgcggtcgg gtccctgaaa 1140
 tcgaacatcg gccacctgga gggagcggca ggcatacgcg gattgattaa agcgggtgctc 1200
 gcattgagtc acgagacgat accgccgagc ttacacgtga gacagctgaa cccgaatatc 1260
 cgggttgagg gaacgtcgtc cgacattgtg aaggaagtcc ggccgtggcc cgcgggttcg 1320
 agacgaaggc ttgcgggcgt cagcgcgttt ggttggtccg gcacgaacgc gcatgtcgtt 1380
 cttgaagaag cggcgccgac tggtagagcg gaagtgcga gcgggttcca ttcccgacct 1440
 ccgcgcgcgg ctgcgcgggc ggctgtcccc ctccgcggag gggacactgg gggcactccc 1500
 gacattgcag gcaactcccga cactgcagac actcccga ca tgcagacac tcccgaactt 1560
 gcagggactg caggcactgc ggcaactacg ggcattgcag acgcgatgta tgtgcttccg 1620
 ctgtccgcgc atggtgcgga cgaactgctc cgggtggcgc gggcatacgg ggaattgctg 1680
 acagcgtcgc acgcaccgag cctgcgtgat ctttgctaca cggccgcagt ccgcccagc 1740
 catcaccgat gccggctcgc tgtttccggc agaacggctg aagaactggc ggcgcagctc 1800
 caggggatca cgatcccttc ccagcgacgg aagacggtat tcgtcttctc gggacaggga 1860
 tcgcaatgga tcggaatggg gcgcagctgg atggaccgcg aacccttat tcgcgaggcg 1920
 ttggaacgct gcgaggccgc catgcggcct tatgtggact ggtcgtgaa agaagaactg 1980
 gcgaagctcg accgcgtcga ggtcattcag cctgcgtctc tcgcgtgca ggtcgccatc 2040
 gccgattgt ggcgttcctg gggaatcgag ccggatgccg tcatcgggca cagcatggga 2100
 gaggtcgcgc ccgctcatgt cgcgggtgcg ctgacgtgc aggatgcggc gcggatcatt 2160
 tgcagccgca gccggctgtt gagccggatc agcggcctgg gcgggatggc gatggtggag 2220
 ctgccgctcg cggaatgtga ggccgtgctg tcgacttaca cggaacgact atcgcccgcg 2280
 gtgtcgaacg gacccaactc caccgtcatc tccggtgaag tcgaagccct ggccgaggctc 2340

gtcgcgacgc tggagcggcg aggcgtgtct tgccggccgg tgaaagtgga cttcgccgcg 2400
 catagcccgc aagtggaccc attgtgcgac gaactcctgc agtcgctcga cgggattcaa 2460
 ccgcggcccc cgaccatacc tttttactcc acggtgaccg gcgcgacgct ggagaccacc 2520
 agcctcgaca gcacgtactg ggctcgcaat ctgcgatcgc cggttctgtt ctggcagggc 2580
 atccgccatc ttgccgacag cgggcacgat gtctttctcg agatcagccc tcatcccatc 2640
 ctgctgcccc ccacggcgcg caatgcggcg ctggttccgt ctctgcgccg cgaccaggac 2700
 gaacgcgggt ccacgtcac gtcgctgggc gccctctatg aggctgggca cactgtcgca 2760
 tggcggaccg tgtacccttc cggcaattgc gtgcgcctgc ccgggtatcc ctggcagcgt 2820
 cgtcgtttct ggctcgacgc ttcccccgcg cgacacgcga tcacgttggg caatccgctg 2880
 ttgggaaaaa gcgtcgaagc ctcgacgcaa cccggcactt tcttctggga gacggaactc 2940
 agtctcgctt ccgtgccttg gctggcagac catcgctgc agggcgaagt cgtcttgccg 3000
 gctactgctg atctcgatat ggctctggcc ggaacttccg agaccttcgg tgaaagtccg 3060
 tgctgctgag agcatgtgac ttccacacag atgctcattg tgccgcgcga cggcagcatg 3120
 acgttgacgc tggccatcgc ggtcgataga cccgggatgg cgtcgtttcg gatttccagc 3180
 cggcaggcat cgacatgggt cctgcatgct tccggggaca ttcgtcagac gcctgcggat 3240
 gcatcgaccg tcccgcggga ttctgcggag acggtgcagg cccgctgccc cacagtgggtg 3300
 ccggcggcgg agctgtggcg tcagatggcg gagcacggcg tcgagtatgg tccggcttcc 3360
 cgcgcgctcg agcagatctg gagttgtcca ggtgaggcga tcgggcgtct gcgtagctcg 3420
 gaaacgcgtt ccactgcgcc ggcttctcct gatgcatgtc tgcagatcat cgccgcggcg 3480
 tttggctccc ccggtggaac ctggctgccc gccggcatcg accggatgcg ctggctgcat 3540
 cccgcacgtt ccgtggtgtg gacgcatgcg cggctggaag gacctatcgc cgatctgtcg 3600
 ctgctggacg gagagggaca actggtcgcc cgcacgagg gtctgcggct gcagcgcctg 3660
 gatgcgtcgg agcgcacga catgcgcggc tggttgcacg aactgcgctg ggtcgtcag 3720
 ccgcacgcgg ctgcagagcc gccggcggcg cgagcggcgc ggtcatggct cattgtcggc 3780
 gctgtggata gcgcgctcac cgcattggct cgcgctaccg gcaaccgctg gacgcagacc 3840
 tcgccggaaa agctcgatga actccagccg ccgctcgagg aaatcgtgtt tttgctcgag 3900
 cacgaaccct catgcgaccg cattctgcat ctctccaga ccttggggcg caccgcttgg 3960
 cgtcaagcac cgcgcctatg gctggtcacg cgcggcgcgc agccggtcga tggacagatc 4020
 ctgcaagccg gtatcgctca ggcgcttctc tggggtttgg gccggaccgt gcattacgaa 4080
 catccggaac tgaactgcac gctgatcgat ctcgatcccc ccggcggcga agaggaaactc 4140
 ctgcacgaac tgctgacgaa caacggcgag aatcaaactc cctttcgcgg cggcgcgcgt 4200
 tacgtcgcgc gcgtggctcg gcacgaagcg gatatgcaac ccgccatgtt caaggccggc 4260
 gatcgccgct tccggctcga gatcgatgcc cccggagtc tcgaccggct gcgcttgcgg 4320
 gccacatcgc gccgcccccc gcaagccggg gaagtggaga ttgaagtctg cgccgcgggc 4380
 ctgaacttcc tcgacgttct gctcgccctc ggcgttatgc ccgacgatgc gcccggcgcg 4440
 attgccggca gccgcgcctt gggcggcgaa tgctcgggccc gtatcgtggc catggggaaa 4500
 ggcgtaaccg acttttcgcat cggagatgaa gtctgggccc ttgcgccttg cagtttcggt 4560
 cgcttcgtca ccacgcccgc ctccgcggtt gccttgaagc cggccaacat tcccgcgaa 4620
 caggccgcgc cctgcctat cgcgtttctc accgcccatt acgcgctctc gcgagcggcg 4680
 cggctggcgc ccggcgaacg agtccgtgatt cacgctgcca ccggcgggtg gggattggcg 4740
 gcaatccaga tcgcacagcg tgcgggcgcg gagatcttcg ctactgccgg gagtccggaa 4800
 aaacgagcgt atctgcgctc gctgggcate gcgcattgtt cggattcgcg ctcgatggct 4860
 ttcgtggacg acatccgcaa ttggacgaat caagaaggag tagacgtcgt cctgaattcg 4920
 ctttccggcg atctgctgga ggcgagcttc gatctgctgc gcgatcatgg acggttcate 4980
 gagatcggca agcgcgatta ctatgccggc cgcaagctgg ggcttcgccc gttcctgaag 5040
 aacctctcgt acacgctggt cgatttgctc ggcattgtccc tgaagcgcgc ggcattgacc 5100
 cgggagctgc tgcaggagat ggtcgcaaaa ttcaatcgg aaacctggcg gccctggaa 5160
 acgcgagtga cgaccatcac cgaatcgggt gaggcgttcc gcaccatggc gcaggcgcgg 5220
 cacatcggca aaatcgtcat ggcgatgcga gattgcgcca atgcgcccac cgcacccta 5280
 cgctcggcgt tcgatagcga gggaacctac ttgattaccg gcggacttgg cgggctcgg 5340


```

cttaccgctcg cacgctggat gatcggacgc ggcgcgccgc ggctggtgct gctgagccgc 5400
cgcgcgcctt caccgaggt ccagcaagcc atcgcgctca tggacgcaga tgtccggacg 5460
gtgcaggccg atgtttctca gcgcgatgaa ctgcagcgcg tgatctcttc catcgatcga 5520
ttgcgcggcg tgattcatgc cgcagccgtt ctgcacgatg cgctgctact gaaccagacg 5580
gaagcgcatt tccgcaacgt gatggccgcg aaaatcgacg gtgcctggaa cctgcacttg 5640
ctcaccgcgc actgcccgtc cgatcatttc gtgctcttct cctccgctgc aggactgctg 5700
ggcgcgcgcc cccagggaaa ctacgcggcc gcgaacgcct ttcttgacgc gctggcctac 5760
taccggaagg cccaaggcct gccggcgctg agcatcggtt ggggtgctg gtccggaggtc 5820
gggtggtgct ccgcgcagga caatcgcgga tcgcggctgg ctttgcgcgg catggaaaac 5880
ctgacgccgc aacacggcct cgctattctg gaacagctgc tgaacagctc ggcttgccac 5940
gtgcgcgcga tgcccatcaa tgcgcgccag tggcggcagt tctatccaa ggcggcgcgag 6000
tctgcactgt tcgagctttt gcatgacgac gcggcgagcg aagccgatgc gccaaacgcg 6060
ttgcgcgcgc ggctgcaatc ggccgagcct cagaccgcga ggacattgct cgaagaacat 6120
ctacagcagc agctggcgcg cgtgctgcgc atcgactctc aaactatcga tccctgcgc 6180
ccgctgaagg aactcggctt cgattccctc atggccctgg agtttcgcaa ccgtctcgaa 6240
ctcacactgg gtctcacgct cccgcgcacc ctgatttggg gtcatccac gctggccggt 6300
cttgcgccgc acctggcgtc gcaaattggga ctgccgctgg tcgaagcgca ggccgcggct 6360
gctgcggaag gagacagccg cgccatgaaa actgcactca gcgggttgga cgacatgtcg 6420
gaagaagcag ccgtggctgc gctccgagga gcaaggtcgt ga 6462

```

<210> 119

<211> 5088

<212> ADN

<213> bacterie

<400> 119

```

gtgaggga aaattgcgcc catgtcgtcg gtcaaactcg cgctattggc gcggaacatg 60
cggcaaaaca tcgcaggett cgacctggtt cagccgaac ccatcgccat cgtcggcatg 120
gcgtgtcggt ttccgggcgg cggaagaat ccggacgcct tctggacgct gttgaagaac 180
ggtgtcgacg gtgtcaccga ggtgccgcca gaccgctgga actcggacca gtactactcc 240
tccgatcccg atgtccggg caaggcgat gcgcgatat ccgccttct cgaacgcatt 300
gacggtttcg atcggaatt ctteggcatc tcccccgcg aagctctgaa catggatccg 360
cagcagcggc tgcgtctgga agtgtgctgg gaagcggcag aggacgccg catctctccc 420
ggccctctgg cgggcagcgc gaccggcgtc tttgcgggt cctgcgccc ggacttcgga 480
ctgtttcagt acgccgacc tgcccgcatc ggagcttgg cgggttccg cgtggcgcat 540
agcatgttgg ccaatcgcat ctctatctg ctgcacctgc gcggtccgag catggcggtc 600
gatacggcct gtcctccgc gtcgtcgcc gtccatctgg cttgccaaag cctgcgccg 660
cgcaatgcg atcggcgatt cgccggcgga gtgaacttga tcctgactcc cgaggcgatg 720
atcgctttgt cgaaggctcg catgttggcg ccgcagggac gctgcaagac gttcgacgcc 780
gcagccgacg gttatgtgcg cggcgagggc tgccgcatcg tgctgctgaa gcggctctcc 840
gatgcgctgg ccgatggcga tgccatccgt gcagtcatcc gcggctcggc aatcaatcag 900
gacggacgga gcaatggcat cacggcgccg aatctgcagg cgcagaaggc ggtcctgcaa 960
gaggcggtgg ccaacgcgca catcgatcca tcccacgtat cgttgatcga ggcgcagggc 1020
acgggcacgt cgctgggcga tcctatcgag atcgaggccc tgcagtcgg ctacgacgcg 1080
ccgactctg cgcttgtct gctgggttcc gtaaagacca acatcgggca tctggagggc 1140
gcggcgggaa tcgccgggct gatcaaagcc gtactcgccc tgcagcatcg caccattcct 1200
ccgcacctgc attttcgccg gctgaatccg aacatctcac tggacggcag ccggtttcgc 1260
atcgccacgg aatcgtcgcc gtggacgtcg gaaggacggc cgcgtctggc cggcgtcagc 1320
tcgttcggtt ttggagggag caacgcgcac gtcacctcg aagaggcgcc tgcactccct 1380

```

```

ttgccgaagc cggtcacacg cccgcagctt ctcactctgt cggcgcgcac cgacgaagcg 1440
ctcggcgaac tggccggcca cttecgaggag ttccctgcagt cgcacccgaa tgcgttgctg 1500
tccgacgttt gcttcaccag tcagggttggg cgcgacgcat atagtccacg cttggcgatc 1560
accgccgcag atgcggcaga ggctgtagcg gcattggccg cggcgcccgcg gcgcgaagta 1620
tcgttgcgcc ggccggccggc aatcgctttt ctcttcacccg gccaggggcg gcagtacgcc 1680
ggcatgggcy cagagcttta taaaacgcag cctgtttttc gcgacgcgct cgatcggttg 1740
gccgattggc tccgtcccca gctcgatgtt ccgctgaccg ttctcttggt cgagtcgggt 1800
tcgccgttgc acgagacggc gtatacccag ccggcaatgt ttgccctgga atgggctctg 1860
gctcagttct ggctgtcgct cggcgctccg ccggactacg tgctgggcca cagtctcggc 1920
gagtatgttg cggcggtgtg ggccggcgcc tttagcgtgg aggacggcct gcggctggtg 1980
accgccaggg ggcggtggt caatgcgctt cccgcggca aagcggtcat cgttcacgcc 2040
aatccgagcc gcacgcggc gctcgccgcc aagggtggcag tcgccgcac gaatgcggcg 2100
gaccgcaccg tgatctccgg cacggctgca gaaatcgcg aagcgcaaga tgacctgcac 2160
cgcgcggcg tggaacgcg agagctgaac gtatcgcatg cgttccattc gccgctgatg 2220
gatccgattt tggacaagtt cgaagcgctt gcagggtgca tcgctatca gccgctggcg 2280
atcccgctgg tgcgaacgt cagcggagcc gtattgccga aaggcacgac actcgacgcc 2340
cgctactggc ggcgacagtt gcgcgaaacc gtgcagtttg aaagcgcat gcgaaccctg 2400
gcggaaccgc agtgcaagct gtttctggaa atcgggccgc atccacgct caccacgctg 2460
ggcgatatt gtctgccga tgacggcgcg gtctggctgc actccctatc taagggacga 2520
tcggattggt ccgtgctgct ggaaagtctt ggcgccctgt ttaccgggg cgtgaatccc 2580
gactggcgcg gtctctatgc cggggaatca ccagccgcg tcgctgctgc gacgtatccg 2640
tttcagcgtg acaccttcag cctgagacgc gtaccgcga gagagccggc gcgcggcggc 2700
atgttgggag cgcgcctcaa cagcgcgttg ggcatgtca tcttcgaaaa ttcgctaacc 2760
acggagacgc ctctgctcca tgagcacgtg atctacgacg cggtcattgt gcccgcgcc 2820
tggcacgtgt cggcatttct cgaagcggca caggaagtct tcggctccgt tccctgcgc 2880
gtctccgatg tcatgatgcg gcaggcactg gccatccgc cggatacgc ggacacgggt 2940
caagcgattg tcacaccgcg cgaggacggc gaagcaaagg tgcaggctct cagccaggat 3000
ggcgattcgt ggaagctcca cacggcagcc agtctgcgc cggcgactgc cggcgccgtt 3060
catttcgagc tgccggcgca gccttcgaa gtcatctccg gcgatgcgt ctacggcgcg 3120
atgaacgcac gcggcgctga tcttggcccc gccttcagtt ggggtggaaga agtctggcgt 3180
cgcgatggcg aggcgctggg gcgaatgcgt ctgccggtg ctgaggatgg cggaacgct 3240
taccggctgc accccggcct gatcgattct tgttttcaag tattcgagc gacttggccc 3300
gcggagcgtt gccagcccgc cgcatacgtg ccggtcggga tcgaagcgg gcgcttctac 3360
cgctccgcgg caggttctct gcgctgtcat gcgcgtctgc gcccgagctc gagcgccccg 3420
ttcgctgggt atctgacgct ggttgaagag accggcgcg tcacgcgca gttttccgga 3480
ctggctgtaa tgcagcccg tacgctgcaa tccgcacagt cgtggctgca ggatgtgcag 3540
tggcaggagt gcgagcgatc gacaacgttg aagtcgcag gccctggcaa gccggaggac 3600
tggttgctgt gtgcggcgcc agacgatgt gccggtttga tgccgcaaga gctgcgcgtc 3660
gtgtccggcg tcaactctcg ccaggcgctg gaacagaccc agactttggt cggccgcccc 3720
gcgcggctct ggctgatcac gcgcggcgct catcgcatca gtgatgacga tgcgactccc 3780
gtcgatcctt tccaggctcc actgtgggga ctccggcagg cgatcgcgcg cgagcatccc 3840
gagctgtggg gcggcctgat cgacctcggt tgcgacaatg ccgacatcg ccccgccatg 3900
ctgctggatg aaatccgtta tgccggcgac gacaaagcga tcgcattgcy caacggacgc 3960
cgctacgttc gccggctggt gcggcacaag gaaacgtcga agcgcccgcc tgccatttca 4020
gccgacggcg tctatctgat caccggcggt ctccggcgcat taggacgaag ggtggcacgc 4080
cgcttgatcg agcaaggcgc gcgcgctctg gtactggctg gccggcatac ggaggcagtt 4140
gcgatctcg agcaactcgg ggctgcagtc atggttgctg cttgcgatgt gatttccgag 4200
caacagctgg cggcgctgct ggccgacccg cgcacccagc cgctgcgtgg agtcgtgcat 4260
gccgcaggcg tgctcgatga cggggtagtt acagaacaga cgtgggctcg tttcgagaag 4320
gtgctggcgc cgaagctgca ggggtgcctg aatcttcacc agctcactcg ccaccatgcy 4380

```

```

ctcgactttt  tctgactctt  ctcttccgcc  gcttcgctgc  tcggttccgc  cggacagagc  4440
aattactcgg  cggccaacgc  atttctcgac  agccttgccc  acatgcgccg  cgcgcaagga  4500
ctaccggcgc  tgagcatcaa  ttggggacca  tgggcgggcg  aaggcatggc  cgcgcgcata  4560
gcgcggcaag  gcctgccggg  ggtaccgctg  ctgccgccgg  aagtgggtgc  gcgcatactt  4620
ggcgatctgc  tgggcgagac  tgccgctcag  atcgcggtgt  tccaagtctc  cgccgaaaaa  4680
aggcggagcc  cggcgagcga  tcccggcttc  atccagcaac  tcaccgaagc  tgcgccggag  4740
cggcggcagg  aactgctgca  gatgcgcata  cgcaagcagg  ccggcggcgt  gctggcgctc  4800
gatgcgtcca  agacgctcga  cccgcgccgg  ccgctcaagg  aatacggaat  cgattcgctg  4860
atggcgctgg  atctggcgcg  cgccatcgga  gagctgggtg  gcaagagcct  tcccgcgaca  4920
ttgctatacg  accatccgac  cgtcgagaaa  ttggccggcc  atgtcctccg  cgaactcgga  4980
ctcgacgtcc  ccagcgattc  cctcgtcgat  gaagtgcggc  agctgtccga  gcaggagatg  5040
gcggcggtta  tcacggaaac  cttgcaccat  ctgggagagg  aacgatga    5088

```

<210> 120

<211> 4306

<212> ADN

<213> bacterie

<400> 120

```

atgagcgatc  tcactcctct  tcaacaggcg  gtcctggcgc  tcaagcgcac  gcgagcgcgt  60
ctcgacgaac  tggagagcgt  ccacaacgaa  cccatcgcca  tcgtcggcat  ggcttgccgc  120
tttcccggcg  cggactcgcc  ggaagcattt  tggcagctcc  tgcacgatgg  catcgatgcc  180
atcccgcaaa  ttcttgcggg  ccgttgggat  gccgatgcgt  ttacgatcc  cgatcccaac  240
gcgcgggaa  agatgtacac  gcgtctgggc  ggattcctcg  atggtgccgt  cgacggcttc  300
gacgcggcgt  tcttcggaat  cagccgcgc  gaggtcgccg  gtctggatcc  gcagcagcgc  360
ctgctgctcg  aggtggcatg  ggaagctttg  gagcgtgcgg  gtcggccgcc  cgacagtctc  420
gcgggcagcg  acaccggagt  gttcatcggg  atcagcaccg  acgactacag  ccggctgaaa  480
cctaccgatc  cggcgctcat  tgacgcctat  accggtaccg  gaaccgcgtt  cagcactgcc  540
gccggacgga  tctcctatct  gctggggttg  cagggaccga  acttccccgt  cgacacggcg  600
tgctcttctt  cactcgtggc  ggttcatctg  gcgtgccgca  gcttgcatgc  gcgagagtgc  660
agcatggcgc  tggccggcgg  cgtgaacctg  attctggcgc  cggaaagcac  gatctacttc  720
tgccgcctgc  gggccatggc  ggccgatggc  cgttgcaaaa  gtttcgctgc  ctccgccgac  780
ggttacggcc  gcggcgaggg  atgcggaatg  ctggtgctga  agcggctgtc  cgatgcgacg  840
cgtgacggcg  atcgtattct  ggcgctgatt  cgcggatcgg  ccgtcaacca  cggcggccgc  900
agcaacggcc  tcacggcgcc  gaacggtcgg  gcgcaggaag  ccgtgattcg  ggcggcgctc  960
aagaacgccg  gcatggcccc  cgccgatgtc  gattacgtgg  aagcccacgg  aaccgggacg  1020
ccgctgggag  atcccatcga  actgcggggc  atggcagcgg  tgctgggcca  ggggcgtgcc  1080
gtcgattctc  cgttgatcgt  cgggtcgggt  atggcagcgg  tgctgggcca  ggggcgtgcc  1140
gcaggtatcg  ccggcctgat  caagaccatt  ctgcacctgc  agcaccgaga  gattccgccc  1200
catctgcatt  tcaacgcgcc  caaccgcgac  gtactctgga  atgagctgcc  gctaaagata  1260
gccaccgcat  gttcgccatg  gccctccaac  ggccgcccc  gagttgccgg  ggtgagctcg  1320
ttcggaatca  gtggcaccaa  ttcgcacgtc  gtcctcgca  aagcgaagac  gaatgtagaa  1380
gcgaagacga  atgtagaggc  gaagacgaat  gtagaggcga  agacgagtga  agaggtcaag  1440
gcgagtgtag  aggccaaaag  gaatgtggag  gctaaggcta  gtgctagtgt  cccctcctc  1500
gagggggaca  gccgcccgcg  aagcggcgcc  ggggggtcgg  gccggccgcc  cagccgcgag  1560
gaagtgccgg  tcccggatca  actccatgcc  gaagacggcc  gcgaatacct  cctaccgctt  1620
tcggcgcgcc  atccgagggc  tctgcgcgat  ctgcggcgcg  cctatcgcca  tgggcgcttt  1680
cacgtccgc  tctccgcgct  gtgttccgcc  gccagcctga  cgcgcagtca  ctacgaacat  1740
cgcgacgcgt  ttgtggcctc  atccctgccc  gagttcaatc  aattgctcga  ggccttccgg  1800

```

cgcaatgaaa ccaatcgcg cgctcgccacc ggtttcgccc atccccggagt tcgtccgaaa 1860
ctcgcccttca tcttttccgg ccagggcgga cagtaccgc gcatggcgta tcgcctgtat 1920
tccgacgagc ctgtcttccg atcggcgatc gaacgttgcg acgccgcctt ccgcagcttc 1980
gtggaatggc ggcttgcgga cctgctcgcc gacgagtcgg gagcatggct gagccagatc 2040
gatcgcgtag agcctgcgct gttcgccgtt caaatcgcgc tggtcgaact gctgcaatcc 2100
tggggaattc gcccggacgg cgtggccgga cacagcatgg gagaagtggc ggcgccccat 2160
gtcgcaggca ttctcaccct ggaggacggc gcccgcata tctgtcgccg cagccggctg 2220
ttgctcggac ttcgcgggcg gggagcgatg gctctggtcg aactgccgct cgatcggggc 2280
aaggccgtgc tcgctgaacg cggctctact actgtttctg tcgcggccag caacggacca 2340
cgacgacggc tgttctcggg agaccgtgtg gctctcgagc atttgaagga cgacttcgag 2400
aggcgcgggc tcttctgccc gctgattcag gtggatgtcg cttcacacag ctgcgaggtg 2460
gaccgcgtcg agaagcaatt gcgccaggaa ctcgcccgcg ttattgcaaa acgttccgcc 2520
gtgcccgttc tctccacggc tgaaggacag ttgagcacgg gcgagggcgtg cgacgcgtcg 2580
tactgggtag ccaatctgcg acagccagtc cgtttctggg agtcgttgca ggcatggct 2640
ggtgatgagt tcacgcagtt cctggagatc agtcgcgcat ctgtgctgac gccgtcgatc 2700
gaggatagtc tgcggacgct cggcataaac ggactggttc gcccgcgtact gcgcgcgac 2760
gaaccggagc ggcgtgagct gctcgagttg ctgcgcgcgc tctacgtgaa tgggcagcgt 2820
ccggactggc gcgcgctcgc ttggtctccc gacacgcgcc tggatctgcc gacgtatccc 2880
tggcagcgcg agcgttctcg gttcgcgacc tcgacgcggc gaagtctgcc ggcagttggc 2940
ggtcatccgc tgcctgggtc caaggtcgag attgcgtggt cgccggacac acacgtctgg 3000
gagtcggtgc tctctctgga tgcgctgccc tttctcgccg atcaccggct caacgagctt 3060
gtggtgcttc ccggtgccc tlatgtggag atggcgctgg ccgcagccaa ggaagtgttc 3120
gcgggtggct gcagcctgga agagatccgg tttgaacaaa tgctggttgt tccttccgcg 3180
ggcgctcgc gagtgacggc cactactcgag ggacacgcat tccgcattct cagtctggcc 3240
gaaggcgggt ccgattggac cgagcacgcg cgcggcacca tggctgcggc gccggacaag 3300
gtcgcgccca cggtagcct gccacactt ggggatcgca tcgagggcga tgacttctat 3360
gcggccttcg catcgacggg gatgcattac ggcgacacct tccgcggcat cgcggaagt 3420
tggcgcgcg acggcgaggc agtggcgcgga ctgagcgtgc cggatgccgt tcgcgaagca 3480
gagtcgggtt acacgcttca tctgccttg ctgatgctt gtttgcaggt gctgggcgcg 3540
acgcttggcg gcgaaggcag cgcgggtcct tgcgtgctg tcgccatcga acggttgac 3600
tgtttcggca gaccgcggc cgatcttagg gtgcatgcgc ggctgacggg gcggctcgag 3660
ggcgatgtca cctgtgtga tgcggaaggc cagctcatcc tcgaggtcca aggcctgct 3720
gcccaggaac tggagcgcca atccgaatgg ttccacgcta tggaatggga gccgcagctg 3780
ctggccgaga gtccaacggc aacggtgtcg ggtgcatggc tggtcattgc cgatgccggc 3840
ggcatcgag ccgcgggtgg gcgagggctg ggcacaaaca cggttgtgat ttcgggtcgc 3900
gatgccgaga taccggatca gccttaccgg ggcgtcattc actgcgggag cctggatgag 3960
accgaggtg agaccgatcc gtcggctgcg gggggaaccg cctgcgaaga cattttgcgc 4020
atcgttcaag aattcgaggt gggacgcata cagctgacga aacaagcgtc cgacgccgaa 4080
tcgcagcatc cgcgaatctg gctgattacg gcggcggttc atgcggagca tctgcagatg 4140
ccggtggtgc ccgcgcgggc accggtgtgg ggtctgggac gtaccatcgc ggccgagcat 4200
cccagttcg cttgcacctg catcgatctc gacactgccg gtgaagtcga ggtgcaggcg 4260
ctctgcccag agattctcgc ggggagttct gaacgtcagg gcccg 4306

<210> 121

<211> 1537

<212> PRT

<213> bacterie

<400> 121

Leu Gln Cys Pro Glu Ser Ala Val Asp Leu Gln Gln Pro Leu Val Arg
 1 5 10 15
 Met Gly Leu Asp Ser Leu Met Ala Val Gln Leu Arg Asn Arg Ile Asp
 20 25 30
 Thr Asp Leu Arg Val Leu Leu Pro Met Val Arg Phe Leu Asp Gly Pro
 35 40 45
 Ser Val Ala Glu Leu Ala Arg Asp Leu Ser Asp Leu Ser Gly Leu Ser
 50 55 60
 Glu Arg Thr Thr Val Ala Pro Glu Pro Ala Ala Gln Ala Ser Val Pro
 65 70 75 80
 Ala Leu Ser Tyr Pro Leu Ser Ala Gly Gln Gln Ala Leu Trp Phe Ile
 85 90 95
 Tyr Arg Ser Ala Pro Glu Ser Pro Ala Tyr Asn Ile Ala Trp Ile Ala
 100 105 110
 Arg Ala Arg Gly Ala Phe Asp Pro Gln Ala Leu Arg Arg Ser Leu Gln
 115 120 125
 Asp Leu Val Asp Arg His Pro Ala Leu Arg Thr Thr Ile Ala Glu Ser
 130 135 140
 Gly Gly Ala Pro Val Gln Thr Val His Ser Ser Val Pro Val Asp Phe
 145 150 155 160
 Glu Val Ile Pro Cys Ser Pro Asp Asp Glu Ala Val Leu Ile Asp Gly
 165 170 175
 Val Phe His Ala Pro Phe Asn Leu Gly Glu Asn Cys Phe Arg Ser Arg
 180 185 190
 Leu Leu Val Gln Ser Gly Lys Asp Gln Val Leu Ala Ile Val Val His
 195 200 205
 His Ile Leu Ala Asp Phe Trp Ser Leu Leu Val Met Val Asp Glu Leu
 210 215 220
 Arg Ser Ile Tyr Leu Ala Arg Thr Ala Gly Gly Pro Pro Val Ala Pro
 225 230 235 240
 Pro Val Ala Ser Phe Ala Ala Phe Val Arg Trp Gln Asn Glu Leu Leu
 245 250 255
 Ala Gly Thr Glu Gly Glu Arg Leu Trp Asn Tyr Trp Ser Ser Gln Leu
 260 265 270

Ser Gly Gln Leu Pro Val Leu Asn Leu Pro Ser Asp Arg Pro Ser Pro
 275 280 285
 Pro Val Gln Ser Phe Arg Gly Asn Ser His Ser Phe Arg Ile Glu Pro
 290 295 300
 Ala Leu Thr Ala Lys Leu Lys Ala Leu Ala Arg Arg Gln Asn Ala Thr
 305 310 315 320
 Leu His Ala Thr Leu Met Ala Ala Phe Gln Val Leu Leu Ser Arg Trp
 325 330 335
 Thr Ser Gln Glu Glu Ile Leu Thr Gly Thr Leu Thr Asn Gly Arg Thr
 340 345 350
 Gln Pro Glu Phe Ala Asp Leu Val Gly Tyr Phe Val Asn Pro Val Ile
 355 360 365
 Leu Arg Gly Glu Leu Ser Gly Asp Pro Asp Phe Asn Thr Val Leu Ala
 370 375 380
 Arg Ile Arg Gln Thr Leu Leu Gly Ala Ile Glu His Gln Glu Tyr Pro
 385 390 395 400
 Tyr Ala Arg Ile Val Glu Arg Leu Gly Pro Gly Leu Arg Val Leu Phe
 405 410 415
 Val Leu Gln Gln Pro His Arg Ile Pro Glu Ser Val Pro Phe Met Leu
 420 425 430
 Gly Gln Ser Gly Gly Arg Met Ala Trp Gly Ser Leu Thr Leu Glu Ser
 435 440 445
 Leu Ala Met Pro Leu Arg Gln Ser Arg Phe Asp Leu Asp Leu Met Met
 450 455 460
 Val Glu Thr Asp Gly Gly Leu Ser Ala Phe Leu Gln Tyr Asn Thr Asp
 465 470 475 480
 Ile Phe Asp Ala Ala Thr Ile Glu Arg Leu Ser Leu His Phe Ala Val
 485 490 495
 Leu Leu Glu Gly Ile Ala Glu Asn Pro Ala Cys Pro Val Val Asp Leu
 500 505 510
 Pro Leu Leu Thr Thr Arg Glu Arg Ile Gln Leu Leu Glu Glu Trp Asn
 515 520 525
 Ala Thr Ala Ala Glu Phe Pro Ser Gln Cys Val His Glu Leu Phe Glu

530		535		540
Ala Gln Val Glu Leu Thr Pro Asp Ala Ile Ala Leu Ser Phe Gly Glu				
545		550		555
				560
Gln Asn Leu Thr Tyr Arg Glu Leu Asn Gly Ser Ala Asn Arg Ile Ala				
	565		570	575
His Tyr Leu Arg Ser Arg Gly Ala Gly Pro Gly Glu Met Val Gly Ile				
	580		585	590
His Val Thr Arg Ser Leu Glu Thr Val Ala Gly Leu Leu Gly Val Leu				
	595		600	605
Lys Ala Gly Ala Ala Tyr Val Pro Leu Glu Pro Glu Tyr Pro Ala Gln				
	610		615	620
Arg Leu Arg Leu Met Leu Glu Glu Thr Arg Pro Val Val Val Leu Asn				
	625		630	635
				640
Val Thr Glu Ser Glu Val Trp Thr Gln Pro Asp Thr Asn Pro Asn Pro				
	645		650	655
Leu Ala Thr Pro Ala Asp Leu Ala Tyr Val Leu Tyr Thr Ser Gly Ser				
	660		665	670
Thr Gly Arg Pro Lys Gly Val Gln Ile Thr His Gln Ala Val Val Asn				
	675		680	685
Phe Leu Ser Ser Met Arg His Glu Pro Gly Ile Ser Asp Arg Asp Thr				
	690		695	700
Leu Leu Ala Leu Thr Thr Phe Met Phe Asp Ile Ser Ala Leu Glu Ile				
	705		710	715
				720
Phe Leu Pro Leu Ser Ala Gly Ala Arg Val Val Val Ala Asn Gln Glu				
	725		730	735
Thr Ala Val Asp Gly Glu Arg Leu Ala Arg Glu Leu Ala Arg Ser Lys				
	740		745	750
Ala Thr Met Met Gln Ala Thr Pro Ala Thr Trp Arg Leu Leu Leu Ala				
	755		760	765
Ser Gly Trp Pro Gly Asp Arg Arg Leu Thr Ala Leu Cys Gly Gly Glu				
	770		775	780
Ala Leu Pro Arg Asp Leu Ala Asp Arg Leu Leu Gln Arg Thr Ala Ala				
	785		790	795
				800

Leu Trp Asn Leu Tyr Gly Pro Thr Glu Thr Thr Ile Trp Ser Ala Ile
 805 810 815

Gln Arg Val Thr Thr Gly Asp Gly Pro Val Ser Ile Gly Arg Pro Ile
 820 825 830

Ala Asn Thr Gln Leu Tyr Val Leu Asp Asp Arg Met Gln Pro Ala Pro
 835 840 845

Ile Gly Val Ala Gly Glu Leu Tyr Ile Gly Gly Ala Gly Leu Ala Arg
 850 855 860

Gly Tyr Leu Asn Arg Pro Glu Leu Ser Ala Asp Lys Phe Val Ala Asn
 865 870 875 880

Ser Phe Asp Pro His Gly Thr Arg Leu Tyr Arg Thr Gly Asp Leu Ala
 885 890 895

Arg Arg Gln Arg Asp Gly Ala Leu Glu Tyr Leu Gly Arg Ile Asp His
 900 905 910

Gln Val Lys Ile Arg Gly Phe Arg Ile Glu Thr Gly Glu Ile Glu Ala
 915 920 925

Ala Val Arg Ser His Pro Ala Val Arg His Ala Val Val Thr Ala Arg
 930 935 940

Glu Asn Asp Ala Ala Gly Lys Tyr Leu Ala Ala Tyr Ile Val Pro Leu
 945 950 955 960

Ala Asp Gly His Arg Ala Thr Ala Ala Ala Asp Thr Phe His Asp Arg
 965 970 975

Val Glu Ser Glu His Val Thr Gln Trp Gln Ser Val Trp Asp Thr Thr
 980 985 990

Tyr Glu Gln Asn Ala Pro Asn Ala Asp Pro Glu Phe Asn Ile Val Gly
 995 1000 1005

Trp Arg Ser Ser Val Thr Gly Glu Pro Ile Pro Ala Ala Glu Met Arg
 1010 1015 1020

Glu Trp Val Gln Asp Ser Val Asp Arg Ile Leu Ala Ser Arg Pro Arg
 1025 1030 1035 1040

Arg Val Leu Glu Ile Gly Cys Gly Thr Gly Leu Leu Leu Phe Arg Val
 1045 1050 1055

Ala Pro His Cys Ser Glu Tyr Trp Ala Thr Asp Phe Ser Gln Lys Ala
 1060 1065 1070

Leu Asp Tyr Ile Ala Ala His Ala Asp Arg Thr Gly Leu Ala Asn Val
 1075 1080 1085

Arg Thr Phe Arg Gln Ala Ala Asp Asp Ala Cys Glu Ile Asp Ser Arg
 1090 1095 1100

Ser Cys Asp Ala Val Val Leu Asn Ser Val Ile Gln Tyr Phe Pro Gly
 1105 1110 1115 1120

Glu Ala Tyr Leu Arg Arg Val Leu Ala Glu Ala Val Arg Val Val Lys
 1125 1130 1135

Pro Gly Gly Ile Val Phe Val Gly Asp Val Arg Ser Leu Pro Leu Leu
 1140 1145 1150

Glu Thr Phe Tyr Ala Ser Leu Glu Val Gln Arg Ala Pro Ala Ser Leu
 1155 1160 1165

Thr Arg Asn Glu Phe Arg Gln Arg Val Arg Ser Leu Ala Ser Gln Glu
 1170 1175 1180

Glu Glu Leu Val Val Asp Pro Ala Phe Phe Phe Ala Leu Arg Glu Gln
 1185 1190 1195 1200

Ile Pro Glu Ile Gly Arg Ile Glu Ile Leu Pro Arg Arg Gly Arg Ser
 1205 1210 1215

His Asn Glu Leu Thr Arg Phe Arg Tyr Gln Ala Ile Leu His Ile Gly
 1220 1225 1230

Ser Arg Glu Ala Glu Glu Pro Glu Ser Asp Arg Arg Arg Cys Gln Thr
 1235 1240 1245

Ala Ala Glu Ile Arg Arg Val Leu Thr Asp Ala Gln Pro Glu Leu Ala
 1250 1255 1260

Ala Phe Thr Glu Ile Pro Asn Ala Arg Leu Thr Ala Glu Ser Ala Ile
 1265 1270 1275 1280

Val Thr Trp Met Asn Gly Asp Glu Ala Pro Glu Thr Leu Gly Glu Leu
 1285 1290 1295

Arg Asp Arg Leu Arg Gln Thr Ser Pro Ser Gly Val Asp Pro Ala Asp
 1300 1305 1310

Leu Trp Arg Met Asp Glu Asp Leu Pro Tyr Arg Val Ala Ile Asp Trp
 1315 1320 1325

Ser Ser His Gly Pro His Gly Arg Phe Asp Ala Thr Phe Cys Arg Ala

1330	1335	1340
Ala Ala Gly Pro Pro Ala Ser Arg Pro Arg Arg Arg Leu Ala Gly Pro		
1345	1350	1355 1360
Tyr Thr Asn Asp Pro Leu Arg Ala Val Tyr Thr Arg Thr Val Val Pro		
	1365	1370 1375
Gln Leu Arg Thr His Leu Lys Glu Lys Leu Pro Asp Tyr Met Ile Pro		
	1380	1385 1390
Thr Ala Trp Val Val Leu His Glu Met Pro Leu Thr Pro Asn Gly Lys		
	1395	1400 1405
Ile Asp Arg Asn Ala Leu Pro Asp Pro Glu Pro Ser Arg Arg Ala His		
	1410	1415 1420
Ala Glu Ala Phe Thr Pro Pro Glu Thr Pro Val Glu Gln Val Leu Ala		
	1425	1430 1435 1440
His Ile Trp Gly Glu Val Leu Gly Met Asp Gly Ile Gly Val His Asp		
	1445	1450 1455
His Phe Phe Asp Ser Gly Gly His Ser Leu Leu Val Thr Gln Met Ile		
	1460	1465 1470
Ala Arg Val Arg Asp Met Leu His Val Glu Val Pro Phe Arg Thr Val		
	1475	1480 1485
Phe Asn Ala Pro Thr Val Arg Gly Phe Ala Val Ala Ile Gln Asp Gly		
	1490	1495 1500
Val Asp Pro Gly Trp Ala Arg Arg Ala Ala Asp Leu Leu Ile Ala Val		
	1505	1510 1515 1520
Ser Gln Met Ser Asp Val Gln Ile Glu Arg Met Met Ser Ala Ala Gln		
	1525	1530 1535

Asp

<210> 122
 <211> 2766
 <212> PRT
 <213> bacterie

<400> 122
 Met Gln Asn Ser Ser Pro Asn Thr Ile Asp Leu Ser Leu Ala Arg Arg

1	5	10	15
Gln Leu Leu Asp Arg Leu Leu Gln Glu Asn Ser Pro Glu His Arg Ile	20	25	30
Pro Arg Arg Glu Asn Arg Asp Ala Ala Pro Leu Ser Leu Ala Gln Gln	35	40	45
Arg Leu Trp Phe Leu His Gln Leu Asp Pro Asp Ser Pro Ala Tyr Asn	50	55	60
Ile Pro Ile Ala Leu His Ile Arg Gly Pro Leu Asp Ile Arg Val Leu	65	70	75
Leu Arg Ser Leu Glu Ala Val Val Gln Arg His Glu Ser Leu Arg Ser	85	90	95
Cys Ile Gly Gly Val Asp Gly Glu Ala Arg Gln Ser Leu Leu Ala Arg	100	105	110
Val Thr Leu Glu Leu Pro Val Val Gln Ala Asp Gly Ile Ala Glu Ala	115	120	125
Arg Gln Met Ala Leu Arg Asp Ala Gln Ile Pro Phe Asp Leu Arg Lys	130	135	140
Pro Pro Leu Leu Arg Thr Lys Leu Ile Cys Leu Asp Asp Lys Gln Gln	145	150	155
Ile Leu Leu Leu Thr Leu Ser His Ile Ile Ala Asp Ala Trp Ser Val	165	170	175
Glu Thr Phe Val Arg Asp Leu Thr Arg Ser Tyr Glu Ala Phe Val Gln	180	185	190
Gly Arg Pro Ser Pro Leu Met Glu Leu Pro Ile Gln Tyr Gly Asp Trp	195	200	205
Ala Val His Gln Gln Thr Ser Leu Asn Gln Thr Ala Gln Gln Tyr Trp	210	215	220
Lys Lys Gln Leu Ser Gly Thr Leu Pro Phe Leu Asp Leu Pro Thr Asp	225	230	235
Arg Pro Arg Pro Ala Gln Gln Thr Trp Arg Gly Ala Val Glu Thr Thr	245	250	255
Ala Leu Gly Arg Asp Leu Thr Asp Gly Leu His Ala Phe Ala Leu Arg	260	265	270

Glu Gly Ala Thr Val Phe Met Thr Ala Ile Ala Ala Phe Gln Val Leu
 275 280 285
 Leu His Arg Tyr Thr Ala Gln Glu Asp Ile Leu Ile Gly Val Pro Val
 290 295 300
 Ala Gly Arg Thr Gln Arg Glu Thr Glu Gly Leu Val Gly Cys Phe Ala
 305 310 315 320
 Asn Met Ile Val Leu Arg Gly Asp Leu Arg Asp Asp Pro Ser Phe Arg
 325 330 335
 Ser Leu Leu Ala Arg Thr Arg Asp Thr Ala Leu Ser Ala Leu Ser His
 340 345 350
 Gln Asp Phe Pro Phe Glu Arg Leu Val Glu Glu Leu His Pro Pro Arg
 355 360 365
 Asp Leu Ser Arg Ser Pro Val Phe Gln Val Ser Phe Ala Leu Leu Pro
 370 375 380
 Asp Ala Pro Ala Ile Thr Val Met Pro Gly Leu Thr Ile Ser Arg Glu
 385 390 395 400
 Tyr Met His Asn Gly Gly Ser Lys Leu Asp Leu Gly Val Thr Leu Glu
 405 410 415
 Pro Ser Gly Asp Gly Leu Met Ala Ser Ala Glu Tyr Asn Thr Asp Leu
 420 425 430
 Phe Asp Ala Ala Thr Ile Ala Ser Leu Leu Asp Ala Tyr Arg Thr Leu
 435 440 445
 Leu Ala Ser Val Val Thr Asp Pro Asp Val Arg Ile Ser Thr Ala Ala
 450 455 460
 Leu Leu Ser Pro Ala Val Arg Ser Arg Met Leu Glu Gln His Asn Ala
 465 470 475 480
 Thr Arg Arg Asp Ala Gly Pro Asn Gly Cys Ala His Glu Leu Val Glu
 485 490 495
 Ala Gln Ala Glu Arg Thr Pro His Ala Val Ala Val Val Phe Glu Asp
 500 505 510
 His Gln Leu Thr Tyr Ala Glu Leu Asn Ala Arg Ala Asn Arg Leu Ala
 515 520 525
 His Arg Leu Ser Ala Ser Gly Ala Gly Pro Gly Lys Ile Ile Ala Leu
 530 535 540

Ala Met Glu Arg Ser Leu Glu Met Val Ile Ala Leu Leu Ala Ile Leu
 545 550 555 560

Lys Ser Gly Ser Ala Tyr Leu Pro Leu Asp Pro Ala His Pro Lys Asp
 565 570 575

Arg Leu Ala Arg Ile Leu Asp Glu Val Gln Pro His Ala Val Leu Thr
 580 585 590

Gln Glu Ala Val Ala Glu Met Met Ala Met Met Ala Met Met Ala Val
 595 600 605

Ala Val Glu Pro Glu Ala Ala Asn Leu Val Ser Gly Ser Lys Pro Asp
 610 615 620

Asp Leu Ala Tyr Ile Ile Tyr Thr Ser Gly Ser Thr Gly Arg Pro Lys
 625 630 635 640

Gly Val Glu Ile Arg His Ser Ser Leu Val Asn Leu Leu Arg Ser Met
 645 650 655

Gln Arg Glu Pro Gly Leu Thr Ala Ala Asp Gly Leu Val Ala Val Thr
 660 665 670

Thr Val Ser Phe Asp Ile Ala Gly Leu Glu Ile Trp Leu Pro Leu Ile
 675 680 685

Thr Gly Ala Arg Val Ile Val Ala Thr Arg Glu Ile Val Val Asp Gly
 690 695 700

Glu Arg Leu Thr Thr Leu Leu Asp Lys Ser Gly Ala Thr Val Met Gln
 705 710 715 720

Ala Thr Pro Ser Gly Trp Arg Gln Leu Leu Asp Ser Gly Trp Lys Pro
 725 730 735

Gly Lys Gly Phe Arg Val Phe Cys Gly Gly Glu Ala Leu Pro Pro Glu
 740 745 750

Leu Ala Arg Arg Ile Leu Asp Ser Gly Val Glu Leu Trp Asn Leu Tyr
 755 760 765

Gly Pro Thr Glu Thr Thr Ile Trp Ser Ala Val His Lys Thr Gln Arg
 770 775 780

Leu Gly Ala Ser Asp Ser Ile Val Pro Ile Gly His Pro Ile Asp Asn
 785 790 795 800

Thr Gln Leu Tyr Ile Leu Asp Ser Arg Met Glu Pro Val Pro Pro Gly

805	810	815
Val Pro Gly Glu Leu Tyr Ile Gly Gly Ala Gly Leu Ala Arg Gly Tyr		
820	825	830
His Arg Asn Pro Glu Leu Thr Arg Glu Lys Phe Arg Glu Trp Arg Asp		
835	840	845
Arg Gly Arg Ile Tyr Ser Thr Gly Asp Leu Ala Arg Tyr Arg Ser Asp		
850	855	860
Gly Ala Val Glu Cys Leu Gly Arg Val Asp Arg Gln Ile Lys Leu Arg		
865	870	875
Gly Phe Arg Ile Glu Pro Ala Glu Ile Glu Ala Ala Ile Glu Thr His		
885	890	895
Ile Ala Val Lys Gln Ala Ile Thr Val Val Lys Asp Asp Arg Leu Ile		
900	905	910
Ala Tyr Leu Val Pro Ala Thr Gly Asp Val Arg Asp Leu Gln Ser Asp		
915	920	925
Leu Arg Ser Trp Leu Ala Thr Arg Leu Pro Asp Tyr Met Ile Pro Ser		
930	935	940
Ala Phe Val Ser Leu Ser Ser Leu Pro Leu Thr Pro Asn Gly Lys Ile		
945	950	955
Asp Ala Asn Ala Leu Pro Gly Leu Pro Thr Thr Pro Val Ala Ala Arg		
965	970	975
Glu Pro Met Arg Gly Asp Val Val Glu Thr Ile Ala Ser Ile Trp Arg		
980	985	990
Glu Val Leu Arg Val Glu His Val Asp Tyr Arg Gln Asn Phe Phe Asp		
995	1000	1005
Val Gly Gly His Ser Leu Met Leu Thr Arg Val Arg Gly Leu Leu Glu		
1010	1015	1020
Glu Arg Leu Gly Leu Thr Leu Ser Val Val Asp Leu Phe Arg His Thr		
1025	1030	1035
Thr Ile Glu Ser Leu Ala Gly Leu Ala Glu Lys Ser Glu Pro Ala Ala		
1045	1050	1055
Ala Glu Pro Ala Ala Ala Val Ala Glu Asp Arg Ile Ala Val Ile Gly		
1060	1065	1070

Met Ala Gly Arg Phe Pro Gly Ala Arg Asn Val Glu Glu Phe Trp Arg
 1075 1080 1085

Asn Leu Arg Asp Gly Val Asp Ser Ile Ala Arg Leu Ser Pro Glu Asp
 1090 1095 1100

Leu Leu Ala Gly Gly Ile Ser Pro Glu Val Phe Gln Asp Pro Ser Tyr
 1105 1110 1115 1120

Val Pro Ala Lys Gly Leu Leu Asp Gly Ile Glu Phe Phe Asp Ala Ala
 1125 1130 1135

Phe Phe Gly Tyr Ser Pro Arg Glu Ala Glu Ile Met Asp Pro Gln His
 1140 1145 1150

Arg Val Phe Leu Glu Cys Ala Trp Glu Ala Met Glu Asn Ala Gly Tyr
 1155 1160 1165

Ala Ala Arg Ser Tyr Lys Gly Ser Ile Gly Val Phe Ala Gly Cys Gly
 1170 1175 1180

Val Asn Thr Tyr Leu Leu Asn Asn Leu Ala Thr Ala Glu Pro Phe Asp
 1185 1190 1195 1200

Phe Ser Arg Pro Ser Ala Tyr Gln Leu Leu Thr Ala Asn Asp Lys Asp
 1205 1210 1215

Phe Leu Ala Thr Arg Val Ser Tyr Lys Leu Asn Leu Arg Gly Pro Ser
 1220 1225 1230

Leu Thr Val Gln Thr Ala Cys Ser Thr Ser Leu Val Ser Val Val Met
 1235 1240 1245

Ala Cys Glu Ser Leu Gln Arg Gly Ala Ser Asp Ile Ala Leu Ala Gly
 1250 1255 1260

Gly Val Ala Ile Asn Val Pro Gln Ser Val Gly Tyr Leu His Gln Pro
 1265 1270 1275 1280

Gly Met Ile Leu Ser Pro Asp Gly Arg Cys Arg Ala Phe Asp Glu Ser
 1285 1290 1295

Ala Gln Gly Thr Val Pro Gly Asn Gly Ala Gly Val Val Val Leu Lys
 1300 1305 1310

Arg Leu Ser Arg Ala Leu Ala Asp Gly Asp Thr Ile Tyr Ala Val Ile
 1315 1320 1325

Arg Gly Ala Ala Ile Asn Asn Asp Gly Ala Glu Arg Met Gly Phe Thr
 1330 1335 1340

Ala Pro Gly Val Asp Gly Gln Thr Arg Leu Ile Arg Arg Thr Gln Glu
1345 1350 1355 1360

Met Ala Gly Val Lys Pro Glu Ser Ile Gly Tyr Ile Glu Ala His Gly
1365 1370 1375

Thr Ala Thr Pro Leu Gly Asp Pro Val Glu Ile Ala Ala Ile Ala Ala
1380 1385 1390

Asn Phe Pro Lys Asn Gly Ser Gly Asp Val Tyr Ile Gly Ser Val Lys
1395 1400 1405

Thr Asn Ile Gly His Leu Asp Val Ala Ala Gly Val Ala Gly Leu Ile
1410 1415 1420

Lys Thr Val Leu Ala Val His Arg Gly Gln Ile Pro Pro Ser Leu Asn
1425 1430 1435 1440

Phe Gln Arg Pro Asn Pro Arg Ile Asp Phe Ala Asn Thr Pro Phe Arg
1445 1450 1455

Val Ser Thr Arg Leu Leu Asp Trp Pro Ala Gly Lys Thr Pro Arg Arg
1460 1465 1470

Ala Ala Val Ser Ser Phe Gly Ile Gly Gly Thr Asn Ala His Val Ile
1475 1480 1485

Leu Glu Gln Ala Pro Pro Val Thr Pro Ala Ala Ala Ala Pro Glu Arg
1490 1495 1500

Ser Ala His Val Leu Cys Leu Ser Ala Asn Thr Asp Ala Ala Leu Glu
1505 1510 1515 1520

Glu Leu Val Arg Ser Tyr Arg Gly His Met Asp Asn Gln Pro Gly Leu
1525 1530 1535

Ser Phe Gly Asp Val Ala Phe Thr Ala Asn Ala Gly Arg Val His Phe
1540 1545 1550

Pro His Arg Ile Cys Ile Val Ala Arg Ser Ser Asp Glu Ala Arg Gln
1555 1560 1565

Arg Leu Thr Glu Ala Arg Arg Val Arg Ile Ala Gln Thr Arg Pro Lys
1570 1575 1580

Ile Ala Phe Leu Phe Thr Gly Gln Gly Ala Gln Tyr Ala Gly Met Gly
1585 1590 1595 1600

Arg Gln Phe Tyr Glu Ser Gln Pro Val Phe Arg Ala Ala Met Asp Glu

1605	1610	1615
Cys Ala Ala Leu Leu Asn Gly Arg Leu Asp Leu Pro Ala Leu Leu Ala 1620	1625	1630
Asp Asp Ala Leu Leu Asp Ala Thr Ala Gly Ala Gln Pro Ala Leu Phe 1635	1640	1645
Ala Leu Gln Trp Ala Leu Ala Gln Leu Trp Lys Ser Trp Gly Val Thr 1650	1655	1660
Pro Asp Leu Val Met Gly His Ser Val Gly Glu Tyr Ala Ala Ala Cys 1665	1670	1675 1680
Ile Ala Gly Ala Val Ser Leu Pro Asp Ala Leu Gly Leu Val Ala Glu 1685	1690	1695
Arg Gly Arg Leu Met Gln Asn Leu Pro Glu Gly Ala Met Ala Ala Val 1700	1705	1710
Ser Ala Gly Glu Gln Arg Cys Ala Ala Ala Ile Thr Ser Arg Val Ser 1715	1720	1725
Ile Ala Ala Ile Asn Gly Pro Ala Glu Val Val Ile Ser Gly Ala Pro 1730	1735	1740
Gln Asp Ile Glu Ser Ala Leu Ala Thr Leu Arg Ala Glu Gly Ile Lys 1745	1750	1755 1760
Thr Gln Met Leu Ala Val Ala Arg Ala Phe His Ser Ser Ser Met Asp 1765	1770	1775
Pro Ile Leu Ala Asp Leu Gln Arg Arg Ala Ala Ala Ile Ala Trp Arg 1780	1785	1790
Asn Pro Ser Ile Gly Leu Val Ser Asn Leu Thr Gly Lys Leu Ala Gly 1795	1800	1805
Glu Gly Gln Leu Ala Asn Pro Leu Tyr Trp Arg Asp His Ala Arg Asn 1810	1815	1820
Pro Val Arg Phe Ala Asp Gly Ile Gln Thr Leu Lys Asp Glu Gly Cys 1825	1830	1835 1840
Asp Val Phe Leu Glu Ile Gly Pro Lys Pro Val Leu Leu Gly Met Gly 1845	1850	1855
Gln Lys Cys Leu Pro Asp Asp Ala Lys Gln Trp Leu Pro Ser Leu Arg 1860	1865	1870

Lys Gly Arg Asp Glu Trp Glu Thr Ile Leu Ser Ser Val Ala Thr Leu
 1875 1880 1885
 Tyr Gln Gly Gly Phe Asp Ile Asp Trp Gln Glu Phe Asp Arg Pro Tyr
 1890 1895 1900
 Ser Arg Arg Arg Val Ala Leu Pro Ala Tyr Pro Phe Glu Arg Arg Arg
 1905 1910 1915 1920
 His Trp Ile Glu Arg Ser Ser Arg Pro Glu Pro Val Ala Val Ala Ser
 1925 1930 1935
 Gly Leu Val Gly Cys Arg Leu Ser Leu Pro Val Ala Asp Val Ile Phe
 1940 1945 1950
 Glu Ser Lys Leu Ser Thr Ala Ser Pro Leu Leu Ser Asp His Arg Tyr
 1955 1960 1965
 Tyr Gly Ser Val Val Ala Pro Ala Val Tyr Phe Leu Ala Met Ala Leu
 1970 1975 1980
 Glu Ala Ser Ala Glu Val Phe Gly Ala Gly Arg His Thr Leu Glu Asn
 1985 1990 1995 2000
 Val Asn Phe Ala His Pro Leu Ile Leu Ser Ala Glu Arg Asp Thr Ala
 2005 2010 2015
 Val Gln Leu Val Leu Ser Gln Ser Asp Asp Arg His Ala Ser Phe Arg
 2020 2025 2030
 Ile Leu Ser Leu Ser Asp Gly Ser Trp Asn Leu His Ala Ala Gly Asn
 2035 2040 2045
 Ile Ala Ala His Ala Gly Val Ala Pro Val Pro Arg Leu Val Asp Glu
 2050 2055 2060
 Arg Arg Pro Ala Val Asp Gly Asp Thr Tyr Tyr Ser Leu Leu Arg His
 2065 2070 2075 2080
 Leu Glu Ile Glu Leu Gly Pro Ser Tyr Arg Arg Ile Gln Arg Ile His
 2085 2090 2095
 Phe Gly Glu Gln Glu Ala Leu Ala Ala Ile Asp Ser Ala Thr Pro Leu
 2100 2105 2110
 Asn Pro Arg Cys Glu Leu Ala Glu Ala Gly Leu Gln Leu Leu Ser Ala
 2115 2120 2125
 Ala Ala Ser Pro Ala Leu Ala Asp Gly Ala Glu His Pro Ile Phe Ala
 2130 2135 2140

Pro Leu Gly Ile Asp Arg Val Cys Phe Tyr Gly Ser Leu Glu Gly Ala
 2145 2150 2155 2160

Val Trp Gly Ala Ala Gln Ile Leu Arg His Ser Pro Asp Gly Phe Thr
 2165 2170 2175

Gly Glu Ala Gln Leu Leu Asp Ser Glu Gly Cys Val Leu Gly Glu Leu
 2180 2185 2190

Gln Gly Val Ser Phe Arg Arg Val Thr Arg Ala Trp Ala Gln Arg Ser
 2195 2200 2205

Glu Arg Lys Pro Glu Leu Tyr Glu Val Glu Trp Arg Pro Glu Pro Leu
 2210 2215 2220

Arg Gln Pro Ser Arg Thr Leu Gln Pro Gly Ala Trp Leu Ile Leu Ala
 2225 2230 2235 2240

Asp Ser Gly Gly Ala Ala Arg Ala Leu Ala Asp Ala Leu Thr Ala Gln
 2245 2250 2255

Gly Glu Met Cys Val Thr Val Pro Pro Ala Gly Glu Tyr Met Ser Leu
 2260 2265 2270

Val Gly Glu Arg Asp Trp Arg Gly Ile Val Asn Leu Tyr Ser Leu Asp
 2275 2280 2285

Asp Tyr Glu Leu Gly Cys Arg Ser Thr Leu Ala Leu Val Lys Ser Leu
 2290 2295 2300

Lys Ser Gly Pro Arg Leu Trp Leu Val Thr Ala Gly Ala Gln Ala Thr
 2305 2310 2315 2320

Ser Ala Val His Asn Pro Met Gln Ala Ala Leu Trp Gly Phe Gly Arg
 2325 2330 2335

Val Ile Ala Arg Glu His Pro Asp Leu Trp Gly Gly Leu Ile Asp Leu
 2340 2345 2350

Asp Pro Asp Asp Ala His Ala Ser Ala Ala Gly Ala Ala Ala Gln Met
 2355 2360 2365

Arg Asp Phe Asp Gly Glu Asp Gln Ser Ala Trp Arg Ser Asn Arg Arg
 2370 2375 2380

Tyr Val Pro Arg Leu Thr Arg Arg Pro Ser Ala Arg Ala Ala Val Arg
 2385 2390 2395 2400

Leu Val Ser Gly Ala Thr Tyr Leu Ile Thr Gly Gly Leu Gly Ala Leu

2405	2410	2415
Gly Leu Thr Val Ala Lys Trp Met Val Glu His Gly Ala Thr Arg Val		
2420	2425	2430
Val Leu Ala Gly Arg Arg Pro Pro Asn Glu Glu Gln Gln Arg Val Leu		
2435	2440	2445
Gln Gln Ile Gly Ala Thr Ala Glu Thr Val Asp Val Ser Arg Glu Glu		
2450	2455	2460
Glu Val Ala Asp Leu Ile Arg Arg Ile His Thr Glu Thr Ser Pro Leu		
2465	2470	2475
Arg Gly Val Ile His Ala Ala Gly Val Leu Asp Asp Gly Val Leu Leu		
2485	2490	2495
Asn Gln Asp Trp Thr Arg Ile Ala Ser Val Met Ala Pro Lys Ala Glu		
2500	2505	2510
Gly Ala Val His Leu His His His Thr Arg Asp Leu Pro Leu Asp Phe		
2515	2520	2525
Phe Val Leu Phe Ser Ser Ala Ser Ser Leu Leu Gly Pro Ala Gly Gln		
2530	2535	2540
Ala Gly Tyr Ala Ala Ala Asn Ala Val Leu Asp Ala Leu Ala His His		
2545	2550	2555
Arg Arg Gly Leu Gly Leu Pro Ala Thr Ser Ile Asn Trp Gly Arg Trp		
2565	2570	2575
Ser Gly Ala Gly Met Ala Ala Arg Thr Ser Gln Ser Met Ala Gly Val		
2580	2585	2590
Ala Ser Leu Ser Val Asp Glu Gly Leu His Ile Leu Glu Ala Val Leu		
2595	2600	2605
His Glu Cys Pro Ile Gln Ile Ala Ala Leu Pro Ala Gly Ser Ile Thr		
2610	2615	2620
Gly Glu Leu Leu Arg Pro Ala Ala Leu Pro Ser Pro Gln Leu Arg Thr		
2625	2630	2635
Arg Leu Asn Glu Ala Thr Pro Arg Gln Arg Glu Ala Ile Leu Ile Ala		
2645	2650	2655
His Ile Arg Glu Ser Leu Ala Arg Phe Val Gly Ile Ala Thr Ser Thr		
2660	2665	2670

Pro Leu Asp Pro Gln Gln Pro Leu Gly Glu Leu Gly Leu Asp Ser Leu
 2675 2680 2685

Met Ala Ile Glu Leu Arg Asn Ser Leu Ser Gln Ser Leu Gly Gln Pro
 2690 2695 2700

Leu Pro Ala Ser Leu Leu Phe Asp Tyr Pro Ser Leu Asp Ala Ile Val
 2705 2710 2715 2720

Ser Tyr Val Leu His Ala Val Phe Pro Pro Glu Ala Ser Pro Val Glu
 2725 2730 2735

Ala Pro Glu Phe Glu Asn Leu Ala Arg Glu Glu Leu Glu Ala Leu Leu
 2740 2745 2750

Asp Ser Arg Leu Ala Gln Val Asp Gln Trp Leu Glu Thr Gln
 2755 2760 2765

<210> 123

<211> 1763

<212> PRT

<213> bacterie

<400> 123

Met Ser Gly Ser Asp Asp Leu Ser Lys Leu Arg Arg Ala Val Ile Ala
 1 5 10 15

Leu Asp Lys Val Gln Lys Arg Ile Asp Gln Leu Glu Ser Ala Arg Ser
 20 25 30

Glu Pro Ile Ala Leu Ile Gly Ala Gly Cys Arg Phe Pro Gly Ala Ser
 35 40 45

Asn Leu Asp Ala Tyr Trp Ser Leu Leu Arg Glu Gly Arg Ser Ala Val
 50 55 60

Arg Glu Val Pro Pro Asp Arg Trp Asp Ile Asp Ala Tyr Tyr Asp Pro
 65 70 75 80

Asp Pro Gly Ala Thr Gly Arg Met Tyr Thr Arg Tyr Gly Gly Phe Ile
 85 90 95

Asp Gln Val Asp Arg Phe Asp Ala Arg Phe Phe Gly Ile Ala Pro Arg
 100 105 110

Glu Ala Ile Ser Leu Asp Pro Gln Gln Arg Leu Leu Leu Glu Val Thr
 115 120 125

Trp Glu Ala Ile Glu Asn Ala Gly Leu Pro Pro Asp Arg Leu Ala Gly
 130 135 140
 Ser Arg Thr Gly Val Phe Met Gly Ile Phe Ser Asn Asp Tyr Tyr Asn
 145 150 155 160
 Leu Gln Met Arg Gly Gly Asp Ala His Ile Asp Ala Tyr Thr Gly Thr
 165 170 175
 Gly Asn Thr Ala Ser Val Ala Ala Gly Arg Leu Ser Tyr Ile Leu Gly
 180 185 190
 Leu Gln Gly Pro Asn Met Ala Ile Asp Thr Ala Cys Ser Ser Ser Leu
 195 200 205
 Val Ala Val His Leu Ala Cys Gln Ser Leu Arg Ser Gly Glu Ser Asp
 210 215 220
 Leu Ala Leu Ala Gly Gly Val Asn Leu Ile Leu Ser Pro Asp Arg Thr
 225 230 235 240
 Ile Tyr Phe Cys Lys Leu Lys Ala Met Ala Ala Asp Gly Arg Cys Lys
 245 250 255
 Ala Phe Asp Ala Ala Ala Asp Gly Tyr Val Arg Gly Glu Gly Cys Gly
 260 265 270
 Val Val Val Leu Lys Arg Leu Ser Asp Ala Leu Arg Asp Arg Asp Pro
 275 280 285
 Val Met Ala Val Ile Arg Gly Thr Ala Ile Asn Gln Asp Gly Arg Ser
 290 295 300
 Asn Gly Leu Thr Ala Pro Asn Gly Pro Ala Gln Glu Ala Val Ile Arg
 305 310 315 320
 Gln Ala Val Gly Asp Ala Arg Leu Gln Thr Leu Asp Val Ser Tyr Val
 325 330 335
 Glu Ala His Gly Thr Gly Thr Pro Leu Gly Asp Pro Ile Glu Ala Gly
 340 345 350
 Ala Leu Ala Ala Ala Leu Gly Ala Gly Arg Thr Asn Gly Asn Lys Leu
 355 360 365
 Lys Leu Gly Ser Val Lys Thr Asn Phe Gly His Leu Glu Ala Ala Ala
 370 375 380
 Gly Val Ala Ala Leu Ile Lys Val Ala Leu Met Leu Gln Asn Glu Ala
 385 390 395 400

Ile Pro Pro His Leu Asn Leu Thr Thr Pro Ser Pro His Ile Asp Trp
 405 410 415
 Asn Thr Leu Pro Leu Glu Ile Pro Ala Arg Leu Thr Pro Trp Pro Val
 420 425 430
 Ala Pro Gly Gly Arg Arg Val Ala Gly Ile Asn Ser Phe Gly Leu Ser
 435 440 445
 Gly Thr Asn Ala His Val Leu Ile Glu Gln Ala Pro Gln Gln Ala Ala
 450 455 460
 Ser Ser Thr Pro Ala Pro Tyr Leu Leu Pro Leu Ser Ala Arg Ser Pro
 465 470 475 480
 Glu Ala Leu Arg Asp Leu Ala Arg Ala Tyr Arg Asp Val Val Asn Asp
 485 490 495
 Asn Pro Ala Asp Thr Cys Tyr Thr Ala Cys Ala Arg Arg Thr Ser Tyr
 500 505 510
 Glu His Arg Ala Ala Phe Thr Gly Thr Asn Ala Gln Asp Leu Met Ala
 515 520 525
 Gly Leu Asp Ser Phe Leu Ala Gly Asn Pro Asn Arg Asp Thr Ala Thr
 530 535 540
 Gly Phe Val Pro Arg Gly Gln Lys Arg Lys Val Val Phe Val Leu Pro
 545 550 555 560
 Gly Gln Gly Ser Gln Trp Pro Gly Met Gly Arg Asp Leu Met Ala Ser
 565 570 575
 Glu Pro Val Phe Arg Ala Ala Ile Glu Glu Cys Gly Arg Ala Met Gln
 580 585 590
 Pro Tyr Val Asp Trp Ser Leu Thr Gln Glu Leu Gln Gly Pro Leu Asp
 595 600 605
 Arg Ile Asp Val Ile Gln Pro Ala Leu Phe Ala Val Gly Val Ala Leu
 610 615 620
 Ala Gly Leu Trp Arg His Trp Gly Ile Glu Pro Asp Ala Val Ile Gly
 625 630 635 640
 His Ser Met Gly Glu Val Ala Ala Ala His Ile Ala Gly Ala Leu Thr
 645 650 655
 Leu Asp Glu Ala Ala Arg Val Ile Cys Leu Arg Ser Arg Met Leu Ala

660	665	670
Gly Val Arg Gly Gln Gly Glu Met Ala Val Val Glu Leu Ala Leu Asp 675 680 685		
Glu Ala Ile Ala Ala Ile Ala Gly Arg Ser Asp Arg Val Ser Ile Ala 690 695 700		
Ala Ser Asn Ser Pro Arg Ser Thr Val Leu Ser Gly Asp Ser Ala Ala 705 710 715 720		
Leu Gly Glu Leu Leu Arg Glu Leu Glu Ala Lys Asp Val Phe Cys Arg 725 730 735		
Arg Val Lys Val Asp Ile Ala Ser His Ser His Leu Met Asp Ser Val 740 745 750		
Cys Ala Ala Leu Pro Gly Val Val Gly Ala Leu Gln Pro Arg Pro Ala 755 760 765		
Ala Leu Gly Met Tyr Ser Thr Val Thr Gly Ala Ala Ile Ser Gly Glu 770 775 780		
Glu Leu Val Ser Ala Tyr Trp Ala Arg Asn Leu Arg Gln Pro Val Met 785 790 795 800		
Leu Ser Thr Ala Val Ala Ala Ala Ala Ala Gly Gly His Asp Val Phe 805 810 815		
Leu Glu Leu Ser Pro His Pro Leu Leu Val Gln Pro Ile Gln Glu Thr 820 825 830		
Leu Gly Asp Arg Ala Ala Ile Ala Ala Ala Ser Leu Arg Arg Asp Glu 835 840 845		
Asp Gly Asn Leu Ala Leu Arg Arg Thr Leu Gly Ala Leu Leu Thr Asn 850 855 860		
Gly Val Thr Pro Asp Trp Ser Arg Ile Tyr Pro Asn Gly Gly Gln Thr 865 870 875 880		
Arg Arg Leu Pro Asn Tyr Pro Trp Gln Arg Glu Arg Tyr Trp Ile Asp 885 890 895		
Ile Arg Pro Pro Gln Val Glu Ser Gln Ala Leu Pro Gly Arg Arg Ile 900 905 910		
Pro Ser Pro Leu Pro Glu Met Gln Phe Glu Ser Thr Val Glu Ala Lys 915 920 925		

Asp Phe Ala Asp His Arg Leu His Asp Val Ile Val Thr Pro Gly Ala
 930 935 940
 Trp His Leu Ala Met Ala Leu Ala Ala Ala Arg Gln Gly Leu Gly Ala
 945 950 955 960
 Gly Pro His His Val Glu His Val Ser Leu Thr Gly Ala Leu Thr Leu
 965 970 975
 Pro Glu Asn Asp Ala Ala Arg Gln Val Gln Leu Val Leu Arg His Glu
 980 985 990
 Glu Gly Gly Gly Ala Ser Phe Arg Ile Tyr Ser Arg Glu Asp Ser Trp
 995 1000 1005
 Lys Leu His Ser Glu Gly Met Leu Gln Ala Gly Asp Ser Thr Ala Ser
 1010 1015 1020
 Ile Asp Leu Asp Ala Ile Arg Ala Arg Cys Thr Ala Glu Leu Thr Ala
 1025 1030 1035 1040
 Asp Ala Phe Tyr Ser Arg Leu Trp Asp Arg Gly Tyr His Phe Gly Pro
 1045 1050 1055
 Thr Phe Arg Thr Ile Gly Pro Ile Trp Arg Gly Asn Gly Glu Val Leu
 1060 1065 1070
 Cys Arg Val Asp Ile Pro Leu Thr Glu Met Gln Thr Ile Asp Cys Cys
 1075 1080 1085
 Leu Gln Leu Pro Ala Ala Leu Val His His Asp Asp Leu Lys Asp Val
 1090 1095 1100
 His Val Pro Val Gly Leu Asp Arg Phe Ser Leu Ala Glu Val Pro Thr
 1105 1110 1115 1120
 Gly Pro Val Trp Gly Tyr Ala Val Leu Arg Pro Asp Ser Thr Val Asp
 1125 1130 1135
 Val Arg Leu Val Thr Gly Thr Gly Ser Val Val Ala Glu Leu Val Gly
 1140 1145 1150
 Leu Gln Ser Arg Val Ala His Ser Gly Gln Leu Gly Glu Ser Glu Ile
 1155 1160 1165
 Pro Thr Trp Thr Val Gln Trp Thr Ala Ser Val Arg Arg Gly Asp Ala
 1170 1175 1180
 Asn Ala Gly Asn Ala Gly Gly Pro Trp Leu Val Ile Gly Glu Pro Ala
 1185 1190 1195 1200

Ile Ala Glu Thr Leu Gln Lys Arg Gly Gln Thr Cys Arg Thr Ala Asp
 1205 1210 1215

Thr Cys Ser Gly Pro Pro Cys Arg Gln Ile Val Tyr Cys Pro Ser Pro
 1220 1225 1230

Arg Ile Asp Asp Leu Leu Ser Val Leu Arg Ser Ile Val Gln Ala Gly
 1235 1240 1245

Trp Pro Glu Pro Pro Arg Leu Trp Leu Leu Thr Arg Gly Ser Ala Ala
 1250 1255 1260

Val Leu Asn Ser Asp Lys Asp Ile Asp Ile Arg Gln Ala Trp Leu His
 1265 1270 1275 1280

Gly Ile Gly Arg Thr Ile Ala Tyr Glu His Pro Glu Leu Arg Cys Thr
 1285 1290 1295

Leu Val Asp Leu Asp Ala His Ser Asn Asp Cys Gly His Leu Ala Thr
 1300 1305 1310

Leu Met Leu Ser Asn Ile Ala Glu Asp Gln Val Ala Ile Arg Gln Gly
 1315 1320 1325

Thr Val Trp Ala Pro Arg Leu Ser Leu His Lys Ile Pro Ser Ala Pro
 1330 1335 1340

Asp Val Ala Phe Arg Ala Asp Ala Thr Tyr Leu Ile Thr Gly Gly Leu
 1345 1350 1355 1360

Gly Gly Leu Gly Leu Gln Val Ala Gly Trp Leu Ala Ala Ala Gly Ala
 1365 1370 1375

Arg His Leu Val Leu Leu Gly Arg Ser Glu Arg Pro Arg Pro Gln Leu
 1380 1385 1390

Glu Gly Val Asn Val Lys Ile Ile His Ala Asp Val Ala Asp Arg Gln
 1395 1400 1405

Gln Leu Ser Asp Ala Leu Ala Ile Ile Asp Arg Asp Met Pro Pro Leu
 1410 1415 1420

Arg Gly Val Phe His Leu Ala Gly Thr Leu Ala Asp Gly Met Leu Leu
 1425 1430 1435 1440

Asn Leu Thr Thr Glu Arg Phe Glu Ala Ala Met Ala Pro Lys Val Ala
 1445 1450 1455

Gly Ala Trp Asn Leu His Glu Leu Thr Ala Gly Arg Pro Leu Asp His

1460	1465	1470
Phe Val Leu Phe Ser Ser Ala Ser Ala Thr Val Gly Ser Pro Gly Gln		
1475	1480	1485
Gly Asn Tyr Ala Ala Gly Asn Ser Phe Leu Asp Ala Leu Ala His Leu		
1490	1495	1500
Arg Arg Ala Gln Gly Leu Pro Ala Val Ser Ile Ala Trp Gly Pro Trp		
1505	1510	1515
Thr Gln Val Gly Leu Ala Ala Gln Ala Asn Arg Gly Asp Arg Leu Ala		
1525	1530	1535
Ala Arg Gly Ile Ser Val Ile Gln Pro Gln Gln Gly Leu Arg Ala Leu		
1540	1545	1550
Tyr Lys Ala Leu Thr Gln Ile Arg Pro His Val Ala Val Met Asn Phe		
1555	1560	1565
Asp Ile Ala Gln Trp Leu Arg Tyr Tyr Pro Ser Ala Ala Ser Met Ser		
1570	1575	1580
Leu Leu Ala Gly Ile Ala Pro Ala Ala Ala Asp Thr Lys Pro Ala Ala		
1585	1590	1595
Asp Met Arg Ser Glu Leu Leu Ala Val Pro Ala Gly Arg Gln Arg Arg		
1605	1610	1615
Ala Arg Leu Glu Thr Leu Leu Met His Glu Ala Gly His Val Leu Arg		
1620	1625	1630
Phe Asp Pro Ala Lys Leu Asp Gly Arg Ala Thr Leu Gly Asp Leu Gly		
1635	1640	1645
Phe Asp Ser Leu Met Ala Leu Glu Phe Arg Asn Arg Leu Glu Ala Gly		
1650	1655	1660
Leu Arg Val Lys Leu Ser Ala Thr Leu Ile Trp Arg Tyr Pro Thr Phe		
1665	1670	1675
Ser Ala Leu Ala Gln His Leu Ala Asp Lys Leu Gly Leu Pro Leu Glu		
1685	1690	1695
Ser Met Ala Gly Asn Ala Glu Pro Ser Thr Val Ala Ala Val Ala Thr		
1700	1705	1710
Leu Ala Thr Val Gly Thr Ala Ala Gly Glu Asp Arg Ser Pro Ala Ala		
1715	1720	1725

Ala Asp Asp Leu Asp Ala Val Ala Asn Gln Ile Ala Gly Leu Gly Asp
 1730 1735 1740

Lys Glu Ile Glu Ala Leu Leu Lys Gln Lys Phe Ala His Phe Ser Gly
 1745 1750 1755 1760

Ala Ser Glu

<210> 124
 <211> 2153
 <212> PRT
 <213> bacterie.

<400> 124

Met Ser Ser Ile Ser Glu Arg Phe Pro Asn Leu Thr Pro Leu Gln Gln
 1 5 10 15

Ala Tyr Leu Thr Leu Glu His Met Gln Arg Arg Leu Asp Ala Ala Glu
 20 25 30

Arg Asp Ala Arg Glu Pro Ile Ala Ile Val Gly Leu Gly Cys Arg Phe
 35 40 45

Pro Gly Gly Asp Gly Pro Asp Glu Phe Trp Gln Met Leu Arg Ser Gly
 50 55 60

Val Asp Ala Ile Arg Glu Val Pro Pro Gly Arg Trp Asp Glu Glu Ser
 65 70 75 80

Val Arg Arg Ile Leu Lys Ser Leu Asn Pro Ala Thr Pro Val Lys Ile
 85 90 95

Gln Ala Gly Phe Leu Asp Ser Ile Asp Gly Phe Asp Asn Asp Phe Phe
 100 105 110

Gly Ile Ser Pro Arg Glu Ala Val Ser Ile Asp Pro Gln Gln Arg Leu
 115 120 125

Leu Leu Glu Val Ala Trp Glu Ala Leu Glu Asp Ala Gly Gln Thr Met
 130 135 140

Glu Gly Leu Ser Gly Ser Arg Thr Gly Val Phe Val Gly Ile His Ser
 145 150 155 160

Gln Ser Ser Asp Tyr Phe Trp Met Gln Thr Ala Asp Gly Ala Arg Ile
 165 170 175

Asp Pro Tyr Thr Ala Thr Gly Thr Ala His Ser Val Ile Ala Gly Arg
 180 185 190
 Leu Ser Tyr Leu Leu Asn Leu Gln Gly Pro Ser Ile Ala Leu Asp Thr
 195 200 205
 Ala Cys Ser Ser Ser Leu Ala Ala Val His Leu Ala Cys Gln Ser Leu
 210 215 220
 Arg Ser Gly Glu Cys Thr Leu Ala Val Ala Gly Gly Val Asn Leu Arg
 225 230 235 240
 Phe Ser Pro Glu Phe Met Tyr Ala Thr Ser Lys Met Gly Thr Ala Ser
 245 250 255
 Pro Ser Gly Arg Cys Arg Ala Phe Asp Ala Ala Ala Asp Gly Ile Val
 260 265 270
 Phe Gly Glu Gly Cys Gly Val Val Val Leu Lys Arg Leu Ser Asp Ala
 275 280 285
 Leu Ala Ala Gly Asp Arg Val Trp Ala Val Val Arg Gly Ser Ala Val
 290 295 300
 Asn Gln Asp Gly Arg Ser Ala Gly Leu Thr Ala Pro Asn Val Val Ser
 305 310 315 320
 Gln Gln Val Val Ile Arg Ser Ala Leu Ala Asn Ala Gly Val Ala Ala
 325 330 335
 Gln Gln Ile Gly Tyr Ile Glu Ala His Gly Thr Gly Thr Pro Leu Gly
 340 345 350
 Asp Pro Ile Glu Ile Glu Ala Leu Ala Glu Thr Val Gly Leu Pro Arg
 355 360 365
 Pro Val Gly Asp Val Cys Ala Val Gly Ser Leu Lys Ser Asn Ile Gly
 370 375 380
 His Leu Glu Gly Ala Ala Gly Ile Ala Gly Leu Ile Lys Ala Val Leu
 385 390 395 400
 Ala Leu Ser His Glu Thr Ile Pro Pro Ser Leu His Val Arg Gln Leu
 405 410 415
 Asn Pro Asn Ile Arg Leu Glu Gly Thr Ser Leu Asp Ile Val Lys Glu
 420 425 430
 Val Arg Pro Trp Pro Ala Gly Ser Arg Arg Arg Phe Ala Gly Val Ser
 435 440 445

Ala Phe Gly Trp Ser Gly Thr Asn Ala His Val Val Leu Glu Glu Ala
 450 455 460

Ala Pro Thr Gly Arg Gly Glu Ala Ala Ser Gly Phe His Ser Arg Pro
 465 470 475 480

Pro Ala Ala Ala Ala Arg Ala Ala Val Pro Leu Ala Glu Gly Asp Thr
 485 490 495

Gly Gly Thr Pro Asp Ile Ala Gly Thr Pro Asp Thr Ala Asp Thr Pro
 500 505 510

Asp Thr Ala Asp Thr Pro Asp Ile Ala Gly Thr Ala Gly Thr Ala Ala
 515 520 525

Thr Thr Gly Ile Ala Asp Ala Met Tyr Val Leu Pro Leu Ser Ala His
 530 535 540

Gly Ala Asp Glu Leu Arg Arg Val Ala Arg Ala Tyr Gly Glu Leu Leu
 545 550 555 560

Thr Ala Ser His Ala Pro Ser Leu Arg Asp Leu Cys Tyr Thr Ala Ala
 565 570 575

Val Arg Arg Thr His His Arg Cys Arg Leu Ala Val Ser Gly Arg Thr
 580 585 590

Ala Glu Glu Leu Ala Ala Gln Leu Gln Gly Ile Thr Ile Pro Ser Gln
 595 600 605

Arg Arg Lys Thr Val Phe Val Phe Ser Gly Gln Gly Ser Gln Trp Ile
 610 615 620

Gly Met Gly Arg Ser Trp Met Asp Arg Glu Pro Val Ile Arg Glu Ala
 625 630 635 640

Leu Glu Arg Cys Glu Ala Ala Met Arg Pro Tyr Val Asp Trp Ser Leu
 645 650 655

Lys Glu Glu Leu Ala Lys Leu Asp Arg Val Glu Val Ile Gln Pro Ala
 660 665 670

Leu Phe Ala Leu Gln Val Ala Ile Ala Ala Leu Trp Arg Ser Trp Gly
 675 680 685

Ile Glu Pro Asp Ala Val Ile Gly His Ser Met Gly Glu Val Ala Ala
 690 695 700

Ala His Val Ala Gly Ala Leu Thr Leu Gln Asp Ala Ala Arg Ile Ile

705		710		715		720
Cys Ser Arg Ser Arg	Leu Leu Ser Arg	Ile Ser Gly	Leu Gly Gly	Met		
	725		730		735	
Ala Met Val Glu	Leu Pro Leu Ala	Glu Cys Glu	Ala Val Leu	Ser Thr		
	740		745		750	
Tyr Thr Glu Arg	Leu Ser Pro Ala	Val Ser Asn	Gly Pro Asn	Ser Thr		
	755		760		765	
Val Ile Ser Gly	Glu Val Glu	Ala Leu Ala	Glu Val Val	Ala Thr	Leu	
	770		775		780	
Glu Arg Arg Gly	Val Ser Cys	Arg Pro Val	Lys Val Asp	Phe Ala	Ala	
785		790		795		800
His Ser Pro Gln	Val Asp Pro	Leu Cys Asp	Glu Leu Leu	Gln Ser	Leu	
	805		810		815	
Asp Gly Ile Gln	Pro Arg Pro	Ala Thr Ile	Pro Phe Tyr	Ser Thr	Val	
	820		825		830	
Thr Gly Ala Thr	Leu Glu Thr	Thr Ser Leu	Asp Ser Thr	Tyr Trp	Ala	
	835		840		845	
Arg Asn Leu Arg	Ser Pro Val	Leu Phe Trp	Gln Gly Ile	Arg His	Leu	
	850		855		860	
Ala Asp Ser Gly	His Asp Val	Phe Leu Glu	Ile Ser Pro	His Pro	Ile	
865		870		875		880
Leu Leu Pro Ala	Ile Gly Gly	Asn Ala Ala	Leu Val Pro	Ser Leu	Arg	
	885		890		895	
Arg Asp Gln Asp	Glu Arg Gly	Ser Met Leu	Thr Ser Leu	Gly Ala	Leu	
	900		905		910	
Tyr Glu Ala Gly	His Thr Val	Ala Trp Arg	Thr Val Tyr	Pro Ser	Gly	
	915		920		925	
Asn Cys Val Arg	Leu Pro Arg	Tyr Pro Trp	Gln Arg Arg	Arg Phe	Trp	
	930		935		940	
Leu Asp Ala Ser	Pro Ala Arg	His Ala Ile	Thr Leu Gly	Asn Pro	Leu	
945		950		955		960
Leu Gly Lys Arg	Val Glu Ala	Ser Thr Gln	Pro Gly Thr	Phe Phe	Trp	
	965		970		975	

Glu Thr Glu Leu Ser Leu Ala Ser Val Pro Trp Leu Ala Asp His Arg
 980 985 990
 Val Gln Gly Glu Val Val Leu Pro Ala Thr Ala Tyr Leu Asp Met Ala
 995 1000 1005
 Leu Ala Gly Thr Ser Glu Thr Phe Gly Glu Ser Pro Cys Val Leu Glu
 1010 1015 1020
 His Val Thr Phe Thr Gln Met Leu Ile Val Pro Arg Asp Gly Ser Met
 1025 1030 1035 1040
 Thr Leu Gln Leu Ala Ile Ala Val Asp Arg Pro Gly Met Ala Ser Phe
 1045 1050 1055
 Arg Ile Ser Ser Arg Gln Ala Ser Thr Trp Val Leu His Ala Ser Gly
 1060 1065 1070
 Asp Ile Arg Gln Thr Pro Ala Asp Ala Ser Thr Val Pro Pro Asp Ser
 1075 1080 1085
 Ala Glu Thr Val Gln Ala Arg Cys Pro Thr Val Val Pro Ala Ala Glu
 1090 1095 1100
 Leu Trp Arg Gln Met Ala Glu His Gly Val Glu Tyr Gly Pro Ala Phe
 1105 1110 1115 1120
 Arg Ala Leu Glu Gln Ile Trp Ser Cys Pro Gly Glu Ala Ile Gly Arg
 1125 1130 1135
 Leu Arg Ser Ser Glu Thr Arg Ser Thr Ala Pro Ala Phe Leu Asp Ala
 1140 1145 1150
 Cys Leu Gln Ile Ile Ala Ala Ala Phe Gly Pro Ala Gly Gly Thr Trp
 1155 1160 1165
 Leu Pro Ala Gly Ile Asp Arg Met Arg Trp Leu His Pro Ala Arg Ser
 1170 1175 1180
 Val Val Trp Thr His Ala Arg Leu Glu Gly Pro Ile Ala Asp Leu Ser
 1185 1190 1195 1200
 Leu Leu Asp Gly Glu Gly Gln Leu Val Ala Arg Ile Glu Gly Leu Arg
 1205 1210 1215
 Leu Gln Arg Leu Asp Ala Ser Glu Arg Ile Asp Met Arg Gly Trp Leu
 1220 1225 1230
 His Glu Leu Arg Trp Val Ala Gln Pro His Ala Ala Ala Glu Pro Pro
 1235 1240 1245

Ala Ala Arg Ala Ala Arg Ser Trp Leu Ile Val Gly Ala Val Asp Ser
 1250 1255 1260

Ala Leu Thr Ala Trp Leu Arg Ala Thr Gly Asn Arg Val Thr Gln Thr
 1265 1270 1275 1280

Ser Pro Glu Lys Leu Asp Glu Leu Gln Pro Pro Leu Glu Glu Ile Val
 1285 1290 1295

Phe Leu Leu Glu His Glu Pro Ser Cys Asp Arg Ile Leu His Leu Leu
 1300 1305 1310

Gln Thr Leu Gly Arg Thr Pro Trp Arg Gln Ala Pro Arg Leu Trp Leu
 1315 1320 1325

Val Thr Arg Gly Ala Gln Pro Val Asp Gly Gln Ile Leu Gln Ala Gly
 1330 1335 1340

Ile Ala Gln Ala Pro Phe Trp Gly Leu Gly Arg Thr Val His Tyr Glu
 1345 1350 1355 1360

His Pro Glu Leu Asn Cys Thr Leu Ile Asp Leu Asp Pro Ala Gly Gly
 1365 1370 1375

Glu Glu Glu Leu Leu His Glu Leu Leu Thr Asn Asn Gly Glu Asn Gln
 1380 1385 1390

Ile Ala Phe Arg Gly Gly Ala Arg Tyr Val Ala Arg Val Ala Arg His
 1395 1400 1405

Glu Ala Asp Met Gln Pro Ala Met Phe Lys Ala Gly Asp Arg Pro Phe
 1410 1415 1420

Arg Leu Glu Ile Asp Ala Pro Gly Val Leu Asp Arg Leu Arg Leu Arg
 1425 1430 1435 1440

Ala Thr Ser Arg Arg Pro Pro Gln Ala Gly Glu Val Glu Ile Glu Val
 1445 1450 1455

Cys Ala Ala Gly Leu Asn Phe Leu Asp Val Leu Leu Ala Leu Gly Val
 1460 1465 1470

Met Pro Asp Asp Ala Pro Gly Ala Ile Ala Gly Ser Pro Arg Leu Gly
 1475 1480 1485

Gly Glu Cys Ser Gly Arg Ile Val Ala Met Gly Lys Gly Val Thr Asp
 1490 1495 1500

Phe Arg Ile Gly Asp Glu Val Val Ala Leu Ala Pro Cys Ser Phe Gly

1505	1510	1515	1520
Arg Phe Val Thr Thr Pro Ala Phe Arg Val Ala Leu Lys Pro Ala Asn	1525	1530	1535
Ile Pro Ala Glu Gln Ala Ala Ala Leu Pro Ile Ala Phe Leu Thr Ala	1540	1545	1550
Asp Tyr Ala Leu Ser Arg Ala Ala Arg Leu Ala Pro Gly Glu Arg Val	1555	1560	1565
Leu Ile His Ala Ala Thr Gly Gly Val Gly Leu Ala Ala Ile Gln Ile	1570	1575	1580
Ala Gln Arg Ala Gly Ala Glu Ile Phe Ala Thr Ala Gly Ser Pro Glu	1585	1590	1595
Lys Arg Ala Tyr Leu Arg Ser Leu Gly Ile Ala His Val Ser Asp Ser	1605	1610	1615
Arg Ser Met Ala Phe Val Asp Asp Ile Arg Asn Trp Thr Asn Gln Glu	1620	1625	1630
Gly Val Asp Val Val Leu Asn Ser Leu Ser Gly Asp Leu Leu Glu Ala	1635	1640	1645
Ser Phe Asp Leu Leu Arg Asp His Gly Arg Phe Ile Glu Ile Gly Lys	1650	1655	1660
Arg Asp Tyr Tyr Ala Gly Arg Lys Leu Gly Leu Arg Pro Phe Leu Lys	1665	1670	1675
Asn Leu Ser Tyr Thr Leu Val Asp Leu Leu Gly Met Ser Leu Lys Arg	1685	1690	1695
Pro Ala Leu Thr Arg Glu Leu Leu Gln Glu Met Val Ala Lys Phe Glu	1700	1705	1710
Ser Glu Thr Trp Arg Pro Leu Glu Thr Arg Val Thr Thr Ile Thr Glu	1715	1720	1725
Ser Val Glu Ala Phe Arg Thr Met Ala Gln Ala Arg His Ile Gly Lys	1730	1735	1740
Ile Val Met Ala Met Arg Asp Cys Ala Asn Ala Pro Ile Ala Pro Leu	1745	1750	1755
Arg Ser Ala Phe Asp Ser Glu Gly Thr Tyr Leu Ile Thr Gly Gly Leu	1765	1770	1775

Gly Gly Leu Gly Leu Thr Val Ala Arg Trp Met Ile Gly Arg Gly Ala
 1780 1785 1790

Arg Arg Leu Val Leu Leu Ser Arg Arg Ala Pro Ser Pro Glu Val Gln
 1795 1800 1805

Gln Ala Ile Ala Val Met Asp Ala Asp Val Arg Thr Val Gln Ala Asp
 1810 1815 1820

Val Ser Gln Arg Asp Glu Leu Glu Arg Val Ile Ser Ser Ile Asp Arg
 1825 1830 1835 1840

Leu Arg Gly Val Ile His Ala Ala Ala Val Leu Asp Asp Ala Leu Leu
 1845 1850 1855

Leu Asn Gln Thr Glu Ala His Phe Arg Asn Val Met Ala Ala Lys Ile
 1860 1865 1870

Asp Gly Ala Trp Asn Leu His Leu Leu Thr Arg Asp Cys Pro Leu Asp
 1875 1880 1885

His Phe Val Leu Phe Ser Ser Ala Ala Gly Leu Leu Gly Ala Pro Ala
 1890 1895 1900

Gln Gly Asn Tyr Ala Ala Ala Asn Ala Phe Leu Asp Ala Leu Ala Tyr
 1905 1910 1915 1920

Tyr Arg Lys Ala Gln Gly Leu Pro Ala Leu Ser Ile Gly Trp Gly Ala
 1925 1930 1935

Trp Ser Glu Val Gly Leu Ala Ala Ala Gln Asp Asn Arg Gly Ser Arg
 1940 1945 1950

Leu Ala Leu Arg Gly Met Glu Asn Leu Thr Pro Gln His Gly Leu Ala
 1955 1960 1965

Ile Leu Glu Gln Leu Leu Asn Ser Ser Ala Cys His Val Ala Ala Met
 1970 1975 1980

Pro Ile Asn Val Arg Gln Trp Arg Gln Phe Tyr Pro Lys Ala Ala Gln
 1985 1990 1995 2000

Ser Ala Leu Phe Glu Leu Leu His Asp Asp Ala Ala Ser Glu Ala Asp
 2005 2010 2015

Ala Pro Asn Ala Leu Arg Ala Arg Leu Gln Ser Ala Glu Pro Gln Thr
 2020 2025 2030

Arg Arg Thr Leu Leu Glu Glu His Leu Gln Gln Gln Leu Ala Arg Val
 2035 2040 2045

Leu Arg Ile Asp Ser Gln Thr Ile Asp Pro Leu Arg Pro Leu Lys Glu
 2050 2055 2060

Leu Gly Phe Asp Ser Leu Met Ala Leu Glu Phe Arg Asn Arg Leu Glu
 2065 2070 2075 2080

Leu Thr Leu Gly Leu Thr Leu Pro Ala Thr Leu Ile Trp Gly His Pro
 2085 2090 2095

Thr Leu Ala Gly Leu Ala Pro His Leu Ala Ser Gln Met Gly Leu Pro
 2100 2105 2110

Leu Val Glu Ala Gln Ala Ala Ala Ala Glu Gly Asp Ser Arg Ala
 2115 2120 2125

Met Lys Thr Ala Leu Ser Gly Leu Asp Asp Met Ser Glu Glu Ala Ala
 2130 2135 2140

Val Ala Ala Leu Arg Gly Ala Arg Ser
 2145 2150

<210> 125

<211> 1695

<212> PRT

<213> bacterie

<400> 125

Met Arg Glu Lys Ile Ala Pro Met Ser Ser Val Lys Leu Ala Leu Leu
 1 5 10 15

Ala Arg Asn Met Arg Gln Asn Ile Ala Gly Phe Asp Leu Val His Ala
 20 25 30

Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Phe Pro Gly Gly Ala
 35 40 45

Lys Asn Pro Asp Ala Phe Trp Thr Leu Leu Lys Asn Gly Val Asp Gly
 50 55 60

Val Thr Glu Val Pro Pro Asp Arg Trp Asn Ser Asp Gln Tyr Tyr Ser
 65 70 75 80

Ser Asp Pro Asp Ala Pro Gly Lys Ala Tyr Ala Arg Tyr Ala Ala Phe
 85 90 95

Leu Glu Arg Ile Asp Gly Phe Asp Ala Glu Phe Phe Gly Ile Ser Pro
 100 105 110

Arg Glu Ala Leu Asn Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Val
 115 120 125
 Cys Trp Glu Ala Ala Glu Asp Ala Gly Ile Ser Pro Gly Pro Leu Ala
 130 135 140
 Gly Ser Ala Thr Gly Val Phe Ala Gly Ser Cys Ala Gln Asp Phe Gly
 145 150 155 160
 Leu Phe Gln Tyr Ala Asp Pro Ala Arg Ile Gly Ala Trp Ser Gly Ser
 165 170 175
 Gly Val Ala His Ser Met Leu Ala Asn Arg Ile Ser Tyr Leu Leu Asp
 180 185 190
 Leu Arg Gly Pro Ser Met Ala Val Asp Thr Ala Cys Ser Ser Ala Leu
 195 200 205
 Val Ala Val His Leu Ala Cys Gln Ser Leu Arg Arg Arg Glu Cys Asp
 210 215 220
 Ala Ala Phe Ala Gly Gly Val Asn Leu Ile Leu Thr Pro Glu Gly Met
 225 230 235 240
 Ile Ala Leu Ser Lys Ala Arg Met Leu Ala Pro Asp Gly Arg Cys Lys
 245 250 255
 Thr Phe Asp Ala Ala Ala Asp Gly Tyr Val Arg Gly Glu Gly Cys Gly
 260 265 270
 Ile Val Leu Leu Lys Arg Leu Ser Asp Ala Leu Ala Asp Gly Asp Ala
 275 280 285
 Ile Arg Ala Val Ile Arg Gly Ser Ala Ile Asn Gln Asp Gly Arg Ser
 290 295 300
 Asn Gly Ile Thr Ala Pro Asn Leu Gln Ala Gln Lys Ala Val Leu Gln
 305 310 315 320
 Glu Ala Val Ala Asn Ala His Ile Asp Pro Ser His Val Ser Leu Ile
 325 330 335
 Glu Ala His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Ile Glu
 340 345 350
 Ala Leu Gln Ser Val Tyr Asp Ala Pro Asp Ser Ala Pro Cys Leu Leu
 355 360 365
 Gly Ser Val Lys Thr Asn Ile Gly His Leu Glu Gly Ala Ala Gly Ile

370	375	380	
Ala Gly Leu Ile Lys	Ala Val Leu Ala Leu	Gln His Arg Thr Ile	Pro
385	390	395	400
Pro His Leu His Phe	Arg Arg Leu Asn	Pro Asn Ile Ser Leu	Asp Gly
	405	410	415
Ser Arg Phe Arg Ile	Ala Thr Glu Ser Ser	Pro Trp Thr Ser	Glu Gly
	420	425	430
Arg Pro Arg Leu Ala	Gly Val Ser Ser Phe	Gly Phe Gly Gly	Ser Asn
	435	440	445
Ala His Val Ile Leu	Glu Glu Ala Pro Ala	Leu Pro Leu Pro	Lys Pro
	450	455	460
Val Thr Arg Pro Gln	Leu Leu Thr Leu Ser	Ala Arg Thr Asp	Glu Ala
	465	470	475
Leu Gly Glu Leu Ala	Gly His Phe Ala Glu	Phe Leu Gln Ser	His Pro
	485	490	495
Asn Ala Leu Leu Ser	Asp Val Cys Phe Thr	Ser Gln Val Gly	Arg Asp
	500	505	510
Ala Tyr Ser His Arg	Leu Ala Ile Thr Ala	Ala Asp Ala Ala	Glu Ala
	515	520	525
Val Ala Ala Leu Ala	Ala Ala Pro Arg Arg	Glu Val Ser Leu	Arg Arg
	530	535	540
Arg Pro Ala Ile Ala	Phe Leu Phe Thr Gly	Gln Gly Ala Gln	Tyr Ala
	545	550	555
Gly Met Gly Ala Glu	Leu Tyr Lys Thr Gln	Pro Val Phe Arg	Asp Ala
	565	570	575
Leu Asp Arg Cys Ala	Asp Trp Leu Arg Pro	Gln Leu Asp Val	Pro Leu
	580	585	590
Thr Val Leu Leu Phe	Glu Ser Val Ser Pro	Leu His Glu Thr	Ala Tyr
	595	600	605
Thr Gln Pro Ala Met	Phe Ala Leu Glu Trp	Ala Leu Ala Gln	Phe Trp
	610	615	620
Leu Ser Leu Gly Val	Arg Pro Asp Tyr Val	Leu Gly His Ser	Leu Gly
	625	630	635
			640

Glu Tyr Val Ala Ala Cys Val Ala Gly Ala Phe Ser Val Glu Asp Gly
 645 650 655
 Leu Arg Leu Val Thr Ala Arg Gly Arg Leu Val Asn Ala Leu Pro Arg
 660 665 670
 Gly Lys Ala Val Ile Val His Ala Asn Pro Ser Arg Ile Ala Ala Leu
 675 680 685
 Ala Ala Lys Val Ala Val Ala Ala Ser Asn Ala Pro Asp Arg Thr Val
 690 695 700
 Ile Ser Gly Thr Ala Ala Glu Ile Ala Glu Ala Gln Asp Asp Leu His
 705 710 715 720
 Arg Ala Gly Val Glu Thr Arg Glu Leu Asn Val Ser His Ala Phe His
 725 730 735
 Ser Pro Leu Met Asp Pro Ile Leu Asp Lys Phe Glu Ala Leu Ala Gly
 740 745 750
 Ala Ile Ala Tyr Gln Pro Leu Ala Ile Pro Leu Val Ser Asn Val Ser
 755 760 765
 Gly Ala Val Leu Pro Lys Gly Thr Thr Leu Asp Ala Arg Tyr Trp Arg
 770 775 780
 Arg Gln Leu Arg Glu Thr Val Gln Phe Glu Ser Ala Met Arg Thr Leu
 785 790 795 800
 Ala Asp Arg Glu Cys Lys Leu Phe Leu Glu Ile Gly Pro His Pro Thr
 805 810 815
 Leu Thr Thr Leu Gly Arg Tyr Cys Leu Pro Asp Asp Gly Ala Val Trp
 820 825 830
 Leu His Ser Leu Ser Lys Gly Arg Ser Asp Trp Ser Val Leu Leu Glu
 835 840 845
 Ser Leu Gly Gly Leu Phe Thr Ala Gly Val Asn Pro Asp Trp Arg Gly
 850 855 860
 Leu Tyr Ala Gly Glu Ser Pro Ser Arg Val Ala Leu Pro Thr Tyr Pro
 865 870 875 880
 Phe Gln Arg Asp Thr Phe Ser Leu Arg Arg Val Pro Ala Arg Glu Pro
 885 890 895
 Ala Arg Gly Gly Met Leu Gly Ala Arg Leu Asn Ser Ala Leu Gly Asp
 900 905 910

Val Ile Phe Glu Asn Ser Leu Thr Thr Glu Thr Pro Leu Leu His Glu
 915 920 925
 His Val Ile Tyr Asp Ala Val Ile Val Pro Gly Ala Trp His Val Ser
 930 935 940
 Ala Phe Leu Glu Ala Ala Gln Glu Val Phe Gly Pro Val Pro Cys Ala
 945 950 955 960
 Val Ser Asp Val Met Met Arg Gln Ala Leu Ala Ile Pro Pro Asp Thr
 965 970 975
 Pro Val Thr Val Gln Ala Ile Val Thr Pro Gly Glu Asp Gly Glu Ala
 980 985 990
 Lys Val Gln Val Phe Ser Gln Asp Gly Asp Ser Trp Lys Leu His Thr
 995 1000 1005
 Ala Ala Ser Leu Arg Ala Ala Thr Ala Gly Ala Val His Phe Glu Leu
 1010 1015 1020
 Pro Ala Gln Pro Ser Glu Val Ile Ser Gly Asp Ala Phe Tyr Gly Ala
 1025 1030 1035 1040
 Met Asn Ala Arg Gly Val Asp Leu Gly Pro Ala Phe Ser Trp Val Glu
 1045 1050 1055
 Glu Val Trp Arg Arg Asp Gly Glu Ala Leu Gly Arg Met Arg Leu Pro
 1060 1065 1070
 Val Ala Glu Asp Gly Ala Asn Ala Tyr Arg Leu His Pro Gly Leu Ile
 1075 1080 1085
 Asp Ser Cys Phe Gln Val Phe Gly Ala Thr Trp Pro Ala Glu Arg Cys
 1090 1095 1100
 Gln Pro Gly Ala Tyr Val Pro Val Gly Ile Glu Ala Val Arg Phe Tyr
 1105 1110 1115 1120
 Arg Pro Pro Ala Gly Ser Leu Arg Cys His Ala Arg Leu Arg Pro Ser
 1125 1130 1135
 Ser Ser Gly Pro Phe Val Gly Asp Leu Thr Leu Val Glu Glu Thr Gly
 1140 1145 1150
 Ala Val Ile Ala Glu Phe Ser Gly Leu Ala Val Met His Ala Gly Thr
 1155 1160 1165
 Leu Gln Ser Ala Gln Ser Trp Leu Gln Asp Val Gln Trp Gln Glu Cys

1170	1175	1180
Glu Arg Ser Thr Thr Leu Lys Ser Asp Gly Pro Gly Lys Pro Glu Asp		
1185	1190	1195 1200
Trp Leu Leu Cys Ala Gly Ala Asp Asp Val Ala Gly Leu Met Pro Gln		
1205	1210	1215
Glu Leu Arg Val Val Ser Gly Val Thr Leu Arg Gln Ala Leu Glu Gln		
1220	1225	1230
Thr Gln Thr Leu Val Gly Arg Pro Ala Arg Leu Trp Leu Ile Thr Arg		
1235	1240	1245
Gly Val His Arg Ile Ser Asp Asp Asp Ala Thr Pro Val Asp Pro Phe		
1250	1255	1260
Gln Ala Pro Leu Trp Gly Leu Gly Gln Ala Ile Ala Arg Glu His Pro		
1265	1270	1275 1280
Glu Leu Trp Gly Gly Leu Ile Asp Leu Gly Cys Asp Asn Ala Asp Ile		
1285	1290	1295
Ala Ala Ala Met Leu Leu Asp Glu Ile Arg Tyr Ala Gly Asp Asp Lys		
1300	1305	1310
Ala Ile Ala Leu Arg Asn Gly Arg Arg Tyr Val Arg Arg Leu Val Arg		
1315	1320	1325
His Lys Glu Thr Ser Lys Arg Pro Pro Ala Ile Ser Ala Asp Gly Val		
1330	1335	1340
Tyr Leu Ile Thr Gly Gly Leu Gly Ala Leu Gly Arg Arg Val Ala Arg		
1345	1350	1355 1360
Arg Leu Ile Glu Gln Gly Ala Arg Arg Leu Val Leu Val Gly Arg His		
1365	1370	1375
Thr Glu Ala Val Ala Asp Leu Glu Gln Leu Gly Ala Ala Val Met Val		
1380	1385	1390
Ala Ala Cys Asp Val Ser Ser Glu Gln Gln Leu Ala Ala Leu Leu Ala		
1395	1400	1405
Asp Pro Arg Thr Gln Pro Leu Arg Gly Val Val His Ala Ala Gly Val		
1410	1415	1420
Leu Asp Asp Gly Val Val Thr Glu Gln Thr Trp Ala Arg Phe Glu Lys		
1425	1430	1435 1440

Val Leu Ala Pro Lys Leu Gln Gly Ala Trp Asn Leu His Gln Leu Thr
1445 1450 1455

Arg His His Ala Leu Asp Phe Phe Val Leu Phe Ser Ser Ala Ala Ser
1460 1465 1470

Leu Leu Gly Ser Ala Gly Gln Ser Asn Tyr Ser Ala Ala Asn Ala Phe
1475 1480 1485

Leu Asp Ser Leu Ala His Met Arg Arg Ala Gln Gly Leu Pro Ala Leu
1490 1495 1500

Ser Ile Asn Trp Gly Pro Trp Ala Gly Glu Gly Met Ala Ala Arg Ile
1505 1510 1515 1520

Ala Arg Gln Gly Leu Pro Gly Val Pro Leu Leu Pro Pro Glu Val Gly
1525 1530 1535

Ala Arg Ile Phe Gly Asp Leu Leu Gly Glu Thr Ala Ala Gln Ile Ala
1540 1545 1550

Val Phe Gln Val Ser Ala Glu Lys Arg Arg Ser Pro Ala Ser Asp Pro
1555 1560 1565

Gly Phe Ile Gln Gln Leu Thr Glu Ala Ala Pro Glu Arg Arg Gln Glu
1570 1575 1580

Leu Leu Gln Met Arg Ile Arg Lys Gln Ala Gly Gly Val Leu Ala Leu
1585 1590 1595 1600

Asp Ala Ser Lys Thr Leu Asp Pro Arg Arg Pro Leu Lys Glu Tyr Gly
1605 1610 1615

Leu Asp Ser Leu Met Ala Leu Asp Leu Ala Arg Ala Ile Gly Glu Leu
1620 1625 1630

Val Arg Lys Ser Leu Pro Ala Thr Leu Leu Tyr Asp His Pro Thr Val
1635 1640 1645

Glu Lys Leu Ala Gly His Val Leu Arg Glu Leu Gly Leu Asp Val Pro
1650 1655 1660

Ser Asp Ser Leu Val Asp Glu Val Arg Gln Leu Ser Glu Gln Glu Met
1665 1670 1675 1680

Ala Ala Phe Ile Thr Glu Thr Leu His His Leu Gly Glu Glu Arg
1685 1690 1695

<210> 126
 <211> 1434
 <212> PRT
 <213> bacterie

<400> 126

Met	Ser	Asp	Leu	Thr	Pro	Leu	Gln	Gln	Ala	Val	Leu	Ala	Leu	Lys	Arg
1				5					10					15	
Thr	Arg	Ala	Arg	Leu	Asp	Glu	Leu	Glu	Ser	Val	His	Asn	Glu	Pro	Ile
			20					25					30		
Ala	Ile	Val	Gly	Met	Ala	Cys	Arg	Phe	Pro	Gly	Ala	Asp	Ser	Pro	Glu
		35					40					45			
Ala	Phe	Trp	Gln	Leu	Leu	His	Asp	Gly	Ile	Asp	Ala	Ile	Arg	Glu	Ile
	50					55					60				
Pro	Ala	Gly	Arg	Trp	Asp	Ala	Asp	Ala	Phe	Tyr	Asp	Pro	Asp	Pro	Asn
65					70					75					80
Ala	Pro	Gly	Lys	Met	Tyr	Thr	Arg	Leu	Gly	Gly	Phe	Leu	Asp	Gly	Ala
				85					90					95	
Val	Asp	Gly	Phe	Asp	Ala	Gly	Phe	Phe	Gly	Ile	Thr	Pro	Arg	Glu	Val
		100						105					110		
Ala	Gly	Leu	Asp	Pro	Gln	Gln	Arg	Leu	Leu	Leu	Glu	Val	Ala	Trp	Glu
		115					120					125			
Ala	Leu	Glu	Arg	Ala	Gly	Arg	Pro	Pro	Asp	Ser	Leu	Ala	Gly	Ser	Asp
	130					135					140				
Thr	Gly	Val	Phe	Ile	Gly	Ile	Ser	Thr	Asp	Asp	Tyr	Ser	Arg	Leu	Lys
145					150					155					160
Pro	Thr	Asp	Pro	Ala	Leu	Ile	Asp	Ala	Tyr	Thr	Gly	Thr	Gly	Thr	Ala
				165					170					175	
Phe	Ser	Thr	Ala	Ala	Gly	Arg	Ile	Ser	Tyr	Leu	Leu	Gly	Leu	Gln	Gly
			180					185					190		
Pro	Asn	Phe	Pro	Val	Asp	Thr	Ala	Cys	Ser	Ser	Ser	Leu	Val	Ala	Val
	195						200					205			
His	Leu	Ala	Cys	Arg	Ser	Leu	Gln	Ser	Arg	Glu	Cys	Ser	Met	Ala	Leu
	210					215					220				
Ala	Gly	Gly	Val	Asn	Leu	Ile	Leu	Ala	Pro	Glu	Ser	Thr	Ile	Tyr	Phe
225					230					235					240

Cys	Arg	Leu	Arg	Ala	Met	Ala	Ala	Asp	Gly	Arg	Cys	Lys	Ser	Phe	Ala
				245				250				255			
Ala	Ser	Ala	Asp	Gly	Tyr	Gly	Arg	Gly	Glu	Gly	Cys	Gly	Met	Leu	Val
				260				265				270			
Leu	Lys	Arg	Leu	Ser	Asp	Ala	Thr	Arg	Asp	Gly	Asp	Arg	Ile	Leu	Ala
				275				280				285			
Leu	Ile	Arg	Gly	Ser	Ala	Val	Asn	His	Gly	Gly	Arg	Ser	Asn	Gly	Leu
				290				295				300			
Thr	Ala	Pro	Asn	Gly	Pro	Ala	Gln	Glu	Ala	Val	Ile	Arg	Ala	Ala	Leu
305				310				315				320			
Lys	Asn	Ala	Gly	Met	Ala	Pro	Ala	Asp	Val	Asp	Tyr	Val	Glu	Ala	His
				325				330				335			
Gly	Thr	Gly	Thr	Pro	Leu	Gly	Asp	Pro	Ile	Glu	Leu	Arg	Ala	Met	Ala
				340				345				350			
Ala	Val	Leu	Gly	Glu	Gly	Arg	Ala	Val	Asp	Ser	Pro	Leu	Ile	Val	Gly
				355				360				365			
Ser	Val	Lys	Thr	Asn	Phe	Gly	His	Leu	Glu	Ala	Ala	Ala	Gly	Ile	Ala
370				375				380							
Gly	Leu	Ile	Lys	Thr	Ile	Leu	Ala	Leu	Gln	His	Arg	Glu	Ile	Pro	Pro
385				390				395				400			
His	Leu	His	Phe	Asn	Ala	Pro	Asn	Pro	His	Val	Leu	Trp	Asn	Glu	Leu
				405				410				415			
Pro	Leu	Lys	Ile	Ala	Thr	Ala	Cys	Ser	Pro	Trp	Pro	Ser	Asn	Gly	Arg
				420				425				430			
Pro	Arg	Val	Ala	Gly	Val	Ser	Ser	Phe	Gly	Ile	Ser	Gly	Thr	Asn	Ser
435				440				445							
His	Val	Val	Leu	Ala	Glu	Ala	Lys	Thr	Asn	Val	Glu	Ala	Lys	Thr	Asn
450				455				460							
Val	Glu	Ala	Lys	Thr	Asn	Val	Glu	Ala	Lys	Thr	Ser	Glu	Glu	Val	Lys
465				470				475				480			
Ala	Ser	Val	Glu	Ala	Lys	Gly	Asn	Val	Glu	Ala	Lys	Ala	Ser	Ala	Ser
				485				490				495			
Val	Pro	Leu	Leu	Glu	Gly	Asp	Ser	Arg	Pro	Arg	Ser	Gly	Gly	Gly	Gly

500	505	510
Ser Gly Arg Pro Pro Ser Arg Glu Glu Val Pro Val Pro Asp Gln Leu		
515	520	525
His Ala Glu Asp Gly Arg Glu Tyr Leu Leu Pro Leu Ser Ala Arg His		
530	535	540
Pro Gln Ala Leu Arg Asp Leu Ala Gly Ala Tyr Arg Asp Gly Arg Phe		
545	550	555
His Ala Pro Leu Ser Ala Leu Cys Ser Ala Ala Ser Leu Thr Arg Ser		
565	570	575
His Tyr Glu His Arg Ala Ala Phe Val Ala Ser Ser Leu Pro Glu Phe		
580	585	590
Asn Gln Leu Leu Glu Ala Phe Arg Arg Asn Glu Thr Asn Arg Gly Val		
595	600	605
Ala Thr Gly Phe Ala Asp Pro Gly Val Arg Pro Lys Leu Ala Phe Ile		
610	615	620
Phe Ser Gly Gln Gly Gly Gln Tyr Pro Arg Met Ala Tyr Arg Leu Tyr		
625	630	635
Ser Asp Glu Pro Val Phe Arg Ser Ala Ile Glu Arg Cys Asp Ala Ala		
645	650	655
Phe Arg Ser Phe Val Glu Trp Arg Leu Ala Asp Leu Leu Ala Asp Glu		
660	665	670
Ser Gly Ala Trp Leu Ser Gln Ile Asp Arg Val Gln Pro Ala Leu Phe		
675	680	685
Ala Val Gln Ile Ala Leu Val Glu Leu Leu Gln Ser Trp Gly Ile Arg		
690	695	700
Pro Asp Gly Val Ala Gly His Ser Met Gly Glu Val Ala Ala Ala His		
705	710	715
Val Ala Gly Ile Leu Thr Leu Glu Asp Ala Ala Arg Ile Ile Cys Arg		
725	730	735
Arg Ser Arg Leu Leu Leu Gly Leu Arg Gly Arg Gly Ala Met Ala Leu		
740	745	750
Val Glu Leu Pro Leu Asp Arg Ala Lys Ala Val Leu Ala Glu Arg Gly		
755	760	765

Leu Thr Thr Val Ser Val Ala Ala Ser Asn Gly Pro Arg Ser Thr Val
 770 775 780
 Phe Ser Gly Asp Arg Val Ala Leu Glu His Leu Lys Asp Asp Phe Glu
 785 790 795 800
 Arg Arg Gly Val Phe Cys Arg Leu Ile Gln Val Asp Val Ala Ser His
 805 810 815
 Ser Ser Gln Val Asp Pro Leu Glu Asn Glu Leu Arg Gln Glu Leu Gly
 820 825 830
 Arg Val Ile Ala Lys Arg Ser Ala Val Pro Phe Phe Ser Thr Val Glu
 835 840 845
 Gly Gln Leu Ser Thr Gly Glu Ala Cys Asp Ala Ser Tyr Trp Val Ala
 850 855 860
 Asn Leu Arg Gln Pro Val Arg Phe Trp Glu Ser Leu Gln Ala Met Ala
 865 870 875 880
 Gly Asp Glu Phe Thr Gln Phe Leu Glu Ile Ser Pro His Pro Val Leu
 885 890 895
 Thr Pro Ser Ile Glu Asp Ser Leu Arg Thr Leu Gly Ile Asn Gly Leu
 900 905 910
 Val Arg Pro Val Leu Arg Arg Asp Glu Pro Glu Arg Arg Glu Leu Leu
 915 920 925
 Glu Leu Leu Ala Ala Leu Tyr Val Asn Gly Gln Arg Pro Asp Trp Arg
 930 935 940
 Ala Leu Ala Ser Ser Pro Asp Thr Arg Leu Asp Leu Pro Thr Tyr Pro
 945 950 955 960
 Trp Gln Arg Glu Arg Phe Trp Phe Ala Thr Ser Thr Arg Arg Ser Leu
 965 970 975
 Pro Ala Val Gly Gly His Pro Leu Leu Gly Arg Lys Val Glu Ile Ala
 980 985 990
 Leu Ala Pro Asp Thr His Val Trp Glu Ser Val Leu Ser Leu Asp Ala
 995 1000 1005
 Leu Pro Phe Leu Ala Asp His Arg Leu Asn Glu Leu Val Val Leu Pro
 1010 1015 1020
 Gly Ala Ala Tyr Val Glu Met Ala Leu Ala Ala Ala Lys Glu Val Phe
 1025 1030 1035 1040

Ala Gly Gly Cys Ser Leu Glu Glu Ile Arg Phe Glu Gln Met Leu Val
 1045 1050 1055
 Val Pro Ser Ala Gly Ala Ser Arg Val Gln Val Ile Leu Glu Gly His
 1060 1065 1070
 Ala Phe Arg Ile Ser Ser Leu Ala Glu Gly Gly Ser Asp Trp Thr Glu
 1075 1080 1085
 His Ala Arg Gly Thr Met Ala Ala Ala Pro Asp Lys Val Ala Pro Thr
 1090 1095 1100
 Val Ser Leu Pro Thr Leu Gly Asp Arg Ile Glu Gly Asp Asp Phe Tyr
 1105 1110 1115 1120
 Ala Ala Phe Ala Ser Gln Gly Met His Tyr Gly Asp Thr Phe Arg Gly
 1125 1130 1135
 Ile Ala Glu Val Trp Arg Arg Asp Gly Glu Ala Val Ala Arg Leu Ser
 1140 1145 1150
 Val Pro Asp Ala Val Arg Glu Ala Glu Ser Gly Tyr Thr Leu His Pro
 1155 1160 1165
 Ala Leu Leu Asp Ala Cys Leu Gln Val Leu Gly Ala Thr Leu Gly Gly
 1170 1175 1180
 Glu Gly Ser Ala Gly Pro Cys Val Pro Val Ala Ile Glu Arg Leu His
 1185 1190 1195 1200
 Cys Phe Gly Arg Pro Ala Gly Asp Leu Arg Val His Ala Arg Leu Thr
 1205 1210 1215
 Gly Arg Leu Glu Gly Asp Val Thr Leu Cys Asp Ala Glu Gly His Val
 1220 1225 1230
 Ile Leu Glu Val Gln Gly Leu Arg Ala Gln Glu Leu Glu Arg Gln Ser
 1235 1240 1245
 Glu Trp Phe His Ala Met Glu Trp Glu Pro Gln Leu Leu Ala Glu Ser
 1250 1255 1260
 Pro Thr Ala Thr Val Ser Gly Ala Trp Leu Val Ile Ala Asp Ala Gly
 1265 1270 1275 1280
 Gly Ile Ala Ala Ala Val Ala Arg Gly Leu Gly Thr Asn Thr Val Val
 1285 1290 1295
 Ile Ser Gly Arg Asp Ala Glu Ile Pro Asp Gln Pro Tyr Arg Gly Val

1300	1305	1310
Ile His Cys Gly Ser Leu Asp Glu Thr Glu Asp Glu Thr Asp Pro Ser		
1315	1320	1325
Ala Ala Gly Gly Thr Ala Cys Glu Asp Ile Leu Arg Ile Val Gln Glu		
1330	1335	1340
Phe Gly Val Gly Arg Ile Gln Leu Thr Lys Gln Ala Ser Asp Ala Glu		
1345	1350	1355 1360
Ser Gln His Pro Arg Ile Trp Leu Ile Thr Ala Gly Val His Ala Glu		
1365	1370	1375
His Leu Gln Met Pro Val Val Pro Ala Arg Ala Pro Val Trp Gly Leu		
1380	1385	1390
Gly Arg Thr Ile Ala Ala Glu His Pro Glu Phe Ala Cys Thr Cys Ile		
1395	1400	1405
Asp Leu Asp Thr Ala Gly Glu Val Glu Val Gln Ala Leu Cys Arg Glu		
1410	1415	1420
Ile Leu Ala Gly Ser Ser Glu Arg Gln Gly		
1425	1430	